THESIS

LANGUAGE VARIATION IN ELECTRONIC TEXT-BASED MESSAGES: THE SOCIAL DIFFERENTIATION OF REPRESENTING SPEECH AND ORALITY IN EMC

Submitted by

Chris Howe

Department of English

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Summer 2013

Master's Committee:

Advisor: Gerald Delahunty

Beth Cloven
Frederique Grim

ABSTRACT

LANGUAGE VARIATION IN ELECTRONIC TEXT-BASED MESSAGES: THE SOCIAL DIFFERENTIATION OF REPRESENTING SPEECH AND ORALITY IN EMC

This investigation presents a sociolinguistic interpretation of the extralinguistic factors that play a role in language variation among users of electronically-mediated communication (EMC). EMC scholarship routinely comments on the nature of written language use to represent speech and orality in EMC. Often the use of reduced colloquial variants to represent speech in EMC, such as *gonna*, *ima*, *diz*, etc., is associated with younger users and medium variables. Despite popular perceptions about language use in electronic environments, however, little is known about the extent to which age differentials and stylistic variations shape language use in EMC. Analysis of email and mobile phone text message data gathered from 33 participants between the ages of 18-38 indicates that the decision to represent speech and orality in EMC with the use of colloquial variants is not significantly correlated with younger participants. Instead, analysis of the data from 2,542 electronic text-based message (ETM) transmissions containing 47, 739 words reveals that only logographic features (e.g. acronyms and abbreviations) can be associated with age. The findings indicate that members of the oldest age group have retained the use of logographic features in their electronic encounters from the time when they were the most remarked feature in EMC. As with the dimension of stylistic variation, the level of formality appears to be the greatest indicator of language variation in EMC. The implications suggest that EMC scholarship needs to broaden its view of language use in ETMs regarding the use of informal variants. The practice of representing aspects of speech and orality

may no longer be a phenomenon that can be primarily associated with medium variables or with teenagers and young adults.

TABLE OF CONTENTS

Introduction

The high variability of language use in electronically-mediated communication (EMC) has been widely documented (Anis, 2007; Barnes, 2003; Baron, 2008; Crystal, 2001; Crystal, 2008; Danet & Herring, 2007; Herring, 2001; Palfreyman & Al Khalil, 2007; Squires, 2010; Tagliamonte & Denis, 2008). The language of EMC is often viewed as a "hybrid" form of communication (Herring, 2001; Tagliamonte & Denis, 2008) that mixes elements of orality and writing (Barnes, 2003) by relying on characteristics that correspond to both sides of the "speaking/writing divide" (Crystal, 2001, p. 28). Less explored, however, is a sociolinguistic interpretation of the factors that play a significant role in the variable linguistic behaviors of EMC users. It is the intention of this study to bridge this gap, by investigating how age differentials and stylistic variations influence EMC users to represent features of speech and orality in the written mode.

Since electronic devices have developed to support written human-to-human interaction, scholars have attempted to determine which factors influence language use in the transmission of electronic text-based messages (ETMs) (Barnes, 2003; Baron, 2008, Crystal 2001; Crystal, 2008; Herring, 2001; Squires, 2010; Tagliamonte & Denis, 2008). Early EMC (and computer-mediated communication, or CMC) scholarship foregrounded the constraints of the medium, promoting somewhat deterministic explanations in the analyses of linguistic patterns and features that make use of informal variants considered unacceptable or inappropriate in formal writing (Herring, 2001; Palfreyman & Al Khalil, 2007; Squires, 2010). As Herring (2001) comments, "One important distinction relates to the *synchronicity* of the participation" (italics in original p. 614). Intrigued by the fact that language use in EMC seems to incorporate features characteristic of both speaking and writing – i.e. the messages can be revisable, asynchronous, and text-based

(similar to writing), and/or ephemeral, synchronous, and informal (similar to speech) – medium variables (e.g. synchronicity, message size, keypad designs, etc.) became an easy target in the analytical description of EMC language (Baron, 2008; Squires, 2010; Tagliamonte & Denis, 2008). In other words, some scholars suggest that time constraints of particular messaging systems somehow significantly influence language use in EMC; the more time users have to compose a message, the more likely they are to adhere to standard spellings and orthographies (Crystal, 2001; Herring, 2001).

Yet, it is important to keep in mind that EMC users need to be understood whenever they attempt to break linguistic rules that are traditionally (or prescriptively) associated with standard English; there is no sense in transmitting an ETM that is unintelligible (Crystal, 2008), regardless of medium constraints. So, drawing upon sociolinguistic perspectives which recognize that individuals behave like others linguistically due to shared sets of expectations, values, and knowledge about language use (Coulmas, 2005; Wardhaugh, 2010, p. 118), this investigation will interpret language variation in EMC as an indicator of group membership whenever users interact electronically (Danet & Herring, 2007, p. 7).

First, it seems appropriate to offer some prefatory notes regarding common stereotypes of EMC users and their use of informal variants. There is little disagreement that the mass media has greatly influenced the public's perceptions about the language of EMC, as well as the users themselves (Baron, 2008; Crystal, 2001; Crystal, 2008; Herring, 2001; Squires, 2010). Conventional wisdom suggests that the decision to use informal written variants – e.g. the use of *dunno* instead of *don't know*, or *ima* instead of *I'm going to*, etc. – is a teenager thing; and several surveys have tried to substantiate a general bias in that direction (Crystal, 2008, p. 89). In 2003, for example, the BBC (British Broadcasting Corporation) reported on a survey ("Young

'prefer texting to calls,'" 2003) conducted by the mobile phone insurer CPP (Card Protection Plan) Group, which indicated that over 80% of individuals under the age of 25 preferred to text rather than call. Similarly, countless articles in the popular press have targeted other forms of CMC arguing that the use of Instant Messenger (IM) in particular is increasing among teenagers and leading to a "breakdown in the English language" (O'Connor, 2005, para. 4), since it is "the linguistic ruin of [the] generation" (Axtman, 2002, para. 3). Similarly, writers for *Wired* magazine describe Internet language as "'a whole new fractured language – definitely not as elegant or polished as English used to be'" (quoted in Hale, 1996, p. 9). In brief, the popular press proclaims that electronically-mediated language is more simplified, fractured, and impoverished than traditional forms of written language (Baron, 2008; Crystal, 2001; Crystal, 2008; Herring, 2001; Tagliamonte & Denis, 2008).

In response to such claims, however, linguists argue that, rather than treating the linguistic features as errors caused by carelessness or lack of knowledge of standard writing, the majority of informal variants found in many EMC varieties are in fact deliberate choices to represent features of speech and orality, to express oneself creatively through language, and/or to economize on time and effort during typing (Barnes, 2003; Baron, 2008; Crystal, 2001; Crystal, 2008; Herring, 2001, p. 617). According to Crystal (2001), the discourse that takes place in EMC is best described as a "new species of communication…more than just a hybrid of speech and writing," since it is complete with its own grammar, lexicon, graphology, and usage conditions (p. 48). Others argue that language use in EMC is best perceived as a unique register, filled with a plethora of distinct varieties of language and language use (Squires, 2010). Simply put, ETMs cannot be equated to other forms of non-electronic written texts (Crystal, 2001, p. 48).

The aim of this study is to add to sociolinguistic research in general, and to an understanding of the factors that influence EMC language use in particular, specifically by examining how age and formality correlate with the linguistic variability found in ETMs. I approach this task from several points of view. First, I outline some of the most salient sociolinguistic trends in relation to the concept of a speech community. Building upon the notions that EMC is in itself a social practice (Herring, 2001), and that users often follow a shared set of sociolinguistic norms and stylistic patterns which are originally acquired in offline contexts (Barnes, 2003, p. 91; Danet & Herring, 2007, p. 7), I demonstrate how language use in ETMs is also influenced, at least in part, by social circumstances. In this regard, the trends that are common to most speech communities will inform the first hypothesis about language use in EMC, which proposes that age is a social factor which influences language variation and change in progress (Chambers, 2002; Labov, 1972, Trudgill, 1974; Wardhaugh, 2010). Since stylistic variation has similarly been shown to produce variable linguistic behavior (Coupland, 2007; Labov, 1972; Rickford & Eckert, 2001; Trudgill, 1974; Wardhaugh, 2010), this concept will also be examined to inform the second hypothesis regarding how variable linguistic features are distributed in relation to two levels of formality.

In this study, I examine the frequency of use of informal variants in an EMC corpus of text message and email correspondence gathered from 33 participants. The data were collected in the spring of 2013 by email; a technique that resulted in a corpus of 2,542 ETM transmissions consisting of a total of 47,739 words. Participants' variable linguistic behaviors were measured against age-specific demographic information, as well as against the perceived level of formality assigned to each ETM by the participants themselves. While Fouser, Inoue and Lee (2000) argue that the desire to convey speech and orality in EMC may be universal, "the methods for doing so

remain constrained by the orthographic system and technical capabilities of word-processing" (p. 53). Therefore, a set of sociolinguistic variables, which previous studies have documented as features used by EMC users to represent speech and orality (Fouser et al., 2000), and which have been found to exhibit some structured heterogeneity (variation) (Tagliamonte & Denis, 2008) in different styles and contexts, will be selected for research.

In the following chapters I will provide a brief history of EMC, review sociolinguistic findings as they relate to the concept of a speech community, discuss common linguistic features of EMC that are believed to represent speech, describe the data and methodology, present a discussion of linguistic analysis, and offers limitations and conclusions.

Chapter 1: Electronically-Mediated Communication: History and Media Varieties

The Development of Networking Systems

The various forms of electronically-mediated communication (EMC) have developed at an astounding rate in a very short time. As a result, the recent phenomenon of "interactive networking" (i.e. human-to-human interaction across electronic networking systems) (Herring, 2001, p. 613) has drastically altered the ways in which humans communicate with each other. Originally designed in the 1960s in the United States, "computer networks caught on almost immediately as a means of interpersonal communication;" (Herring, 2001, p. 613). The appeal was originally established among computer scientists in the 1970s, then among associates of elite organizations and universities in the 1980s, before it finally came into popular use in the 1990s through the rise of commercial Internet providers, which facilitated an infrastructure for transmitting written language online (Baron, 2008; Crystal, 2001; Herring, 2001). The rapid development of written language technologies now offers EMC users a wide range of ways to transmit a message to one or more recipients.

By now there is a great amount of scholarship focusing on how to distinguish the various forms of electronic media. One common practice is to differentiate EMC along two dimensions (Barnes, 2003; Baron, 2008; Crystal, 2001; Herring, 2001; Tagliamonte & Denis, 2008). One concerns the audience scope; the transmission of information is intended either for one person (one-to-one) or for a group of recipients (one-to-many) (Barnes, 2003; Baron, 2008; Crystal, 2001; Tagliamonte & Denis, 2008). The other dimension concerns the synchronicity of the communicative event; the communication either occurs with little or no delay (synchronous) or with some delay (asynchronous) (Barnes, 2003; Baron, 2008; Crystal, 2001; Tagliamonte & Denis, 2008). (Note: the difference between synchronous and asynchronous communication is

not a polar opposite distinction; instead, the two forms of communication are separated along a continuum (Baron, 2008, p. 15)). Although medium constraints factor in to these two dimensions and their subdivisions, Herring (2001) rightfully notes that "the deterministic influence of the [electronic] medium on language use is often overstated" (p. 614). Nevertheless, it is helpful to distinguish how one medium is perceived to shape language use differently from another, in order to understand why deterministic explanations of language patterns persist.

Synchronicity

It is often stated that one deliberate practice that influences EMC users to represent speech and orality in ETMs is the choice to "economize on typing effort;" i.e. to save time and energy in particular forms of media believed to be more synchronous (Herring, 2001, p. 617). Research suggests that, in more synchronous modes of EMC, users have a tendency to: (a) omit subject pronouns, auxiliaries, and determiners; (b) use a wide variety of abbreviations; (c) not correct typos; and (d) not use mixed cases (Baron, 2008; Crystal, 2001; Crystal, 2008; Danet & Herring, 2007; Herring, 2001). As a result, research in EMC has typically indicated that synchronicity is one medium variable that seems to have a relatively powerful influence over the structural complexity of the language found in ETMs (Anis, 2007; Baron, 2008; Crystal, 2008; Danet & Herring, 2007; Herring, 2001; Ko, 1996). As with unplanned speech, the relationship between lower levels of structural complexity (e.g. low lexical density, shorter informational units, etc.) and synchronous forms of EMC is often believed to be a reflection of "cognitive constraints on real-time language encoding," (Herring, 2001, p. 617). So while technological constraints cannot fully replicate turn-taking practices characteristic of face-to-face interaction, research maintains that the linguistic complexity of more synchronous modes is highly reduced

relative to more asynchronous ones (Anis, 2007; Baron, 2008; Danet & Herring, 2007; Herring, 2001).

For example, Baron (2004) found that, when she compared 23 IM conversations to Chafe and Danielewicz's (1987 cited in Baron, 2004) contrastive analysis of written and spoken language, average transmission length more closely resembled informal face-to-face interaction. In Baron's study (2004), individual transmissions averaged 5.4 words, "informal spoken conversational intonation units averaged 6.2 words, [and] academic lectures came in at 7.3 words" (p. 57). In contrast, traditional informal letters averaged 8.4 words per punctuation unit, while formal academic papers averaged 9.3 words (Baron, 2008, p. 57). (Note: Baron (2004) describes these averages as "clumps of language set off" by the traditional definition of a standardized punctuation mark that separates structural units in a chunk of text (p. 57)).

In a similar study of language use in a synchronous chat-like EMC protocol called InterChange, Ko (1996), found that users produced shorter words and fewer complements than in a corpus of formal writing that was comparable in size. Findings from both Ko (1996) and Baron (2004) reveal that ETMs tend to have lower average lexical density scores and numbers of words per transmission than either writing or speech in more synchronous modes; a trend which suggests that the burden of producing and processing ETMs in highly synchronous modes does not permit users an extended time for message planning, albeit requiring more conscious attention than talking (Herring, 2001, p. 617).

On the other hand, research suggests that users engaged in more asynchronous modes of communication tend to produce messages with more structural complexity (Baron, 2008; Crystal, 2001; Herring, 2001). Web Logs (aka Blogs), newsgroups, and bulletin boards are perhaps the most popular one-to-many, asynchronous forms of EMC (Baron, 2008), which are

believed to exhibit more complex language structures specifically because people have more time to compose their thoughts, resulting in more standard spelling and orthographic conventions (Barnes, 2003).

In principle, email is also considered to be an asynchronous medium, though people can choose to compose either a one-to-one or a one-to-many email message (Baron, 2008). The popular perception that the linguistic structure of emails is similarly complex has led to presuppositions that, over the past thirty years, email software has increasingly dictated, and subsequently, standardized the linguistic structure for users to adhere to a fixed sequence of discourse elements (e.g. the format of headers, the use of initial terms of endearment, such as *Dear,* closings, etc.) (Crystal, 2001, pp. 94-101). Not to mention the fact that current word-processing and editing applications facilitate the possibilities of revising since the time when these computer functions were limited to considerably archaic types of computer code. Thus, newer applications may have also led email users to expect their interactants to take the time to compose messages with more formal language usage.

Although some users may feel the need to adhere to such prescriptions (and indeed, some do), Baron (2008) argues that EMC actually has "very little tangible data beyond anecdotes [because] researchers are often hesitant to ask colleagues – or strangers – for logs of their email correspondence" (p.16). Instead, the majority of empirical studies which proclaim email to be an asynchronous form of CMC were often documenting language use in other one-to-many public fora, such as bulletin boards and newsgroups, and proposing that similar linguistic trends and patterns occur in private email correspondence (Baron, 2008). In short, proclamations about language use in emails were merely attempts to overgeneralize observations made from different but purportedly comparable forms of asynchronous EMC.

Yet, language use in emails can be distinguished from other forms of asynchronous EMC, specifically because improvements in computer servers and transmission speeds have greatly reduced the lag time in message transmission (sometimes as short as one or two seconds) (Baron, 2008). In other words, technological developments have allowed email to be used as a more synchronous form of communication, if the user wishes to use it in such a manner (Baron, 2008). So, if synchronicity plays as significant a role in language use as commonly presumed, research needs to redefine its understanding of where certain media should be placed along the continuum of synchronicity.

Another form of EMC that echoes this dilemma of synchronicity is the transmission of mobile phone text messages (hereafter text messages). Traditionally, text messaging has been classified as a one-to-one, synchronous medium (Baron, 2008); however, much like email, mobile phones: (a) can transmit text messages that are intended for either one-to-one or one-to-many communication; and (b) allow messages to be stored for recipients to read at their convenience (i.e. interpersonal communication via mobile phones can also be asynchronous by nature). Recall that "In an asynchronous situation, the interactions are stored in some format, and made available to users upon demand, so that they can catch up with the discussion, or add to it, at any time" (Crystal, 2001, p. 11). In fact, transmission speeds on mobile phones approximate more closely to the time lag of emails, since transmission speeds need not take more than one or two seconds, so long as the device is functioning properly.

Another crucial point is that, since traditional mobile phones have advanced, the amount of time needed to tap simple written messages on a numerical keypad has diminished greatly (see Anis, 2007; Baron, 2008; Crystal, 2008; and Herring, 2001 for detailed descriptions of previously documented technological constraints). In 2008, Crystal commented on the

inefficiency of the now outdated keypad layout, which prompted users to develop "several strategies for symbol selection, all of which are awkward and time consuming" (p. 65-66). For example, to send the sequence *cuz*, one needed to (adapted from Crystal, 2008, p. 66):

1. tap the '1' key three times to display a 'c'
2. wait for flashing cursor to complete the time-out process
3. tap the '8' key twice to display a 'u'
4. wait for flashing cursor to complete the time-out process
5. tap the '9' key four times to display a 'z'

The sequence *cuz*, therefore, required nine keystrokes plus two pauses. To Crystal's (2008) relief, however, mobile phone keypads now offer the flexibility of the QWERTY keyboard design (i.e. the most common standard keyboard layout in modern-day use), or a touchscreen equivalent of the QWERTY design. Naturally, it is a consumer choice to purchase a phone equipped with any one of these functions; however, in February of 2010, *Consumer Electronics* reported on a National Purchase Diary (NPD) group consumer survey, which found that all top ten mobile phones in the total volume of retail sales had either QWERTY or touchscreen that displays the same design (Palenchar, 2010). Nokia conducted a similar study in August, 2012, and found that less than 9% of the respondents preferred the old number keypad input method (Ly, 2012). In brief, technological advancements and their appeal are effectively reducing the time and inefficient strategies that are said to make inputting graphic symbols into a message difficult and time consuming.

Clearly, there is a need to inform research about people's actual language use as it pertains to relatively analogous forms of electronic media, rather than formulate assumptions about users' linguistic behaviors by simply extending empirical insights from one mode of communication to other, nonequivalent ones. Furthermore, although research of CMC and EMC has repeatedly demonstrated that electronic text-based language use varies depending on the type

of medium (Baron, 2008; Crystal, 2001; Herring, 2001), literature in this area has neglected to explore the significance of external social and sociocultural factors (Squires, 2010). In the sections that follow, I will examine how certain language variation can be socially marked, in order to make the case that the similar social factors are likely to influence the language practices of EMC users whenever they interact electronically.

Chapter 2: Speech Communities

Chapter Overview

Every branch of linguistics that refers to collective social behavior rests, in part, on the concept of the speech community (Patrick, 2002). Although, the term speech community has been difficult to define (Patrick, 2002; Wardhaugh, 2010), one predominant feature of a speech community is that it refers to a group of speakers who establish some set of normative behaviors based on shared social attitudes, knowledge, and values towards language use (Labov, 1972). Each group's linguistic behaviors relative to other groups are subsequently treated as indicators that mark group membership (Chambers, 2002). Another common feature of a speech community is that differential use of linguistic variation can sometimes project changes to the language (Chambers, 2002; Labov, 1972; Trudgill, 1974; Wardhaugh, 2010). This study will build on the sociolinguistic premise that the linguistic behaviors of EMC users are likely to correspond to ones found in geographically-bound speech communities, assuming they share sociolinguistic norms and values originally acquired in face-to-face contexts (Danet & Herring, 2007, p. 7). In order to make this case, a discussion of the linguistic features and patterns found in EMC will follow this chapter's review of sociolinguistic research, which has shown that linguistic variations are related to, and at times, governed by certain sociological factors.

The Role of Sociological Factors

Sociolinguists generally identify different sociological factors, such as age, sex, and socioeconomic class, as well as levels of formality, when discussing speakers' linguistic behavior in a speech community. Sociolinguistic research has shown that these factors have a tendency to correlate more or less with variable linguistic behaviors (Herring, 2001; Wardhaugh, 2010). Note, however, that correlations between social factors and linguistic variants are not

always associated with language change (Chambers, 2002). This is because group affiliations are limitless and may vary (especially in EMC environments), and because not all members of the same social group always exhibit the same characteristic behavior in every circumstance (Chambers, 2002; Wardhaugh, 2002). The fact that individuals may behave differently linguistically in different situations and circumstances is central to this type of investigation, since language use in EMC is highly variable among individuals and groups, even within a single mode (Herring, 2001). In the sections that follow, I will present a series of studies which observed how a speaker's use of particular linguistic variants have been shown to correlate with one or more social categories, and how certain age cohorts may lead in the vanguard of linguistic change (Chambers, 2002).

Linguistic Variation and Change

First, it is necessary to distinguish between variation and change. All languages have some form of internal variation (Wardhaugh, 2010). Regional variation, for example, may involve the mapping of dialects within regionally identified boundaries (e.g. the dialectal use of *y'all* instead of *you all* due to differences in geographical regions) (Wardhaugh, 2010). From a sociolinguistic point of view, linguistic variation is often defined as the use of a specific set of linguistic units by a particular group of comparatively homogeneous speakers (Wardhaugh, 2010, p. 23), and the use of some identifiable variant is viewed as a "function of external factors, such as sex, age, style, register, and social class" (Antilla, 2002, p. 206). An example can be taken from Chambers (2002) description of Shuy's study (cited in Chambers, 2002), in which it was discovered that certain socially differentiated class- and sex-based patterns influence the use or nonuse of multiple negation in the vernacular of inner-city Detroit African Americans. In Shuy's study, as outlined by Chambers (2002), multiple negation was seen to be used more

frequently by women ranked higher in the social strata, whereas multiple negation in other middle class dialects was relatively nonexistent (Chambers, 2002, p. 353). Similarly, women tended to use multiple negation more frequently than their male counterparts (Chambers, 2002). Thus, the use of multiple negation was seen as a function of external factors such as sex and class among participants in this Detroit study.

In order to explore the relationship between linguistic variation and sociological factors, sociolinguists must identify a sociolinguistic variable. A sociolinguistic variable is simply a unit in the language that has identifiable variants (Wardhaugh, 2010), and that is associated with certain social and/or conventional values. Values of the sociolinguistic variable, therefore, may function as markers of group membership, and the social categories they indexe tend to be distinguished by the social values and attitudes associated with the particular values of the sociolinguistic variable (Chambers, 2002).

Language change, however, occurs when the frequency of use and the values associated with particular competing sociolinguistic variables have become accepted and conventionalized as the norm by members of a speech community (Labov, 1972, p. 2). Historically, language change was viewed as a phenomenon occurring only between different, yet substantially separated successive points in time (i.e. the common practice of investigating language change called diachronic analysis), specifically because linguistic analyses relied heavily on textual evidence that was assumed to be static (Chambers, 2002, p. 356). For example, a common procedure would have been to compare the different forms of language between Shakespearean and contemporary English. In contrast, modern sociolinguists more commonly investigate change as it occurs in apparent time (as well as real time changes, though these are investigated

to a much lesser extent, due to the difficulty in exactly replicating studies with the same pool of participants), in order to map language change in progress.

For Labov (1972), "the problem of explaining language change seems to resolve itself into three separate problems… [and] the model which underlies this three-way division requires" a starting point (p. 1); that is, a point at which there is apparent variation in how one or many speakers produce one or several words differently from other speakers. The first stage begins with the introduction, or origination, of one or many linguistic variations. Typically, the majority of such innovative and incoming variations are believed to be snuffed out almost immediately, though a few survive (Labov, 1972, p. 2). When they do survive, the new forms may be imitated more or less frequently, until their spread confronts older forms (Labov, 1972, p. 2). Eventually, one of the two competing forms triumphs, at which point "regularity has been achieved" (Labov, 1972, p. 2). Regularity (i.e. the notable reoccurrence of the some linguistic variable when produced under relatively fixed conditions or circumstances) in use of a certain sociolinguistic variable is thus seen as a marker of change approaching completion. (Note, however, that some changes may never fully reach completion, such as the Great Vowel Shift, which is still in progress (Wardhaugh, 2010)).

Sociolinguists have traditionally, though not exclusively, tended to focus on phonological variables to study sociological differentiations in speech communities. Yet the rise in new technologies, which supports textually-based human-to-human interaction across electronic devices, implies that the field needs to broaden its scope of theoretical inquiry and include discussions regarding linguistic variations that affect the writing system as well. Prior to making the case that extralinguistic factors such as age and style can inform the sociolinguistic inquiry into language variability in EMC, however, a review of literature relevant to the survey of speech

communities requires attention. So, let us identify some common trends regarding the ways in which speech patterns can be associated with external factors in society.

Sociological Factors and Language Variation

Predictions about change in progress typically stem from analyses of language variation. The field of sociolinguistics has traditionally, though not exclusively, studied speakers' use of phonological variants as a function of external social factors (Anttila, 2002, p. 206). When tracing back to some of the earliest studies investigating language variation, linguists often refer to Fischer's 1958 study (cited in Labov, 1972; Patrick, 2002; Trudgill, 1974; Wardhaugh, 2010). Fischer was interested in understanding how the sociolinguistic variable (ng) differentiated stylistically (Labov, 1972) when subject to varying levels of formality among young children in a New England speech community (Wardhaugh, 2010). The two variants, [n] and [ŋ], have been documented as a stable variation with a long-standing history; and research indicates that [n] is usually the stigmatized form, being produced in contexts that are less formal (Wardhaugh, 2010, p. 166), whereas the velar [ŋ] variant is reserved for the most formal situations (Labov, 1972). Perhaps the most significant findings from this early study of language variation is that the (ng) variable "reflected sensitivity to sex, formality and cultural orientation toward school" (Labov 1972: 238). Though Fischer's prototypical investigation of language variation was not an attempt to make any profound claims, due to the limited number of subjects and highly informal data collection methods (Wardhaugh, 2010, p. 168), external factors (in the case of Fischer's study, sex and formality) came to be reoccurring influences on an individual's linguistic behavior (Chambers, 2002; Labov, 1972; Trudgill, 1974; Wardhaugh, 2010).

In 1962, Labov (1972) expanded the study of language variation and change by looking at the interrelation of sociolinguistic variables in two dimensions: stylistic differentiation and

social stratification. In his "New York department store study," Labov (1972) was interested in observing the differential use of post-vocalic [r] in the New York City (NYC) speech community. To carry out this experiment, Labov (1972) selected three department stores that showed clear, distinct demarcations with regards to the social class groups to which they catered. The linguistic variable was cleverly isolated as the sales associates were asked the whereabouts of an item that was located on the fourth floor. Labov (1972) would then indicate to the sales clerk that he had misunderstood the original response, which prompted the informant to repeat their response. The assumption was that the second utterance would instantiate a more careful repetition of *fourth floor* (Wardhaugh, 2010, p. 168), thereby demonstrating that individuals in the NYC speech community indeed adjust their pronunciation of the (r) variable to adopt the more prestigious form during careful speech.

The findings generated several theoretical issues for sociolinguistic research. Most generally the study found that each subgroup of the NYC speech community demonstrated a differential use of (r) in the same order as the extralinguistic factors (i.e. socioeconomic statuses) that stratified them (Labov 1972). That is, realizations of [r]-use occurred more often among members of the higher social statuses, and less so among members of lower social statuses.

Later, in 1966, Labov (1972) conducted another, more sophisticated study of how linguistic variables interrelated with different styles of speech and social class differentiators in the Lower East Side NYC speech community. Five distinct social classes were identified, which exhibited regular variation of five phonological variables in different contextual styles (Labov 1972). Once the dependent phonological variables had been identified, Labov (1972) developed a series of techniques to isolate five different contextual styles and investigate how the stylistic dimension interrelated with the dependent variables whenever they were elicited in varying

levels of formality (i.e. from the most casual or informal to the most careful or formal types of speech).

Most notably, Labov (1972) discovered that, while members of the NYC speech community exhibited a great range in the absolute values of the variables, there was an identifiable pattern of stylistic variation (Labov, 1972). With regards to the (r) variable, the respondents showed consistent agreement in that [r]-pronunciation in words like *car* and *bird* appeared to be associated with prestige. The prestigious social values associated with the presence of post-vocalic [r] were so prevalent, that members of the lower middle class outperformed members of the upper middle class in the most formal contexts. This "crossover" phenomenon is what Labov (1972) calls hypercorrection. Labov (1972) argued that hypercorrection tends to occur when individuals consciously outperform a group they regard as socially superior in an attempt to approximate to the linguistic behaviors of the higher class in more formal contexts. For Labov (1972), such a phenomenon is indicative of an acceleration of change in progress.

At approximately the same time, we find another study of social differentiation and language variation in British English. In a methodological framework similar to Labov's 1966 study, Trudgill (1974) explored sixteen sociolinguistic variables and their interrelationship with levels of formality and the three prominent social factors (e.g. class, gender, and age) in the city of Norwich, England. Members of the Norwich speech community were assigned to five social indices, and four contextual styles were isolated to study variation of the phonological variables among speakers of the Norwich speech community. For brevity, we shall focus only on Trudgill's (1974) analysis of the (ng) variable.

Data gathered from the Norwich questionnaire indicated that members of the Norwich speech community showed an increase in the use of [ŋ] endings as they moved from everyday speech to more formal styles (Trudgill 1974). As in Fischer's study (cited in Labov, 1972; Trudgill, 1974; Wardhaugh, 2010), differences in [ŋ] values positively correlated with levels of formality and social class. Though speakers of Norwich did not exhibit any crossover effect or hypercorrection, Trudgill (1974) did show a phenomenon that exhibited relatively close consistency with Labov's (1972) findings: while the second highest social index demonstrated the steepest gradient in stylistic variation in the NYC speech community, in Norwich it appeared that members of the third highest social index had the greatest awareness of the social significance of the linguistic variable, due to "the 'border-line' nature of their social class position" (Trudgill, 1974, p. 92). Trudgill (1974) attributed this phenomenon to the linguistic insecurity of the upper working class and notes that the (ng) variable is an important distinguisher of middle class and working class speakers in Norwich.

Although these studies were not conducted in real-time – i.e. the data was not gathered over intervals of time to determine if the findings represent actual changes to the language – follow-up studies have been conducted to verify the findings (Wardhaugh, 2010). In 1983, Trudgill replicated his earlier study with data gathered from seventeen additional participants born between 1958 and 1973 (as cited in Wardhaugh, 2010, p. 210). Analysis of the younger participants' (then aged between 10 and 25) linguistic behavior showed that some of the changes he had originally identified had progressed (Wardhaugh, 2010, p. 210). For example, the words *fin* and *thin* became indistinguishable with [f] being substituted for [θ] more frequently by younger people.

Similarly, Fowler's 1986 replication of Labov's department store study demonstrated higher scores of [r]-use among all classes, styles and age groups; an indication that a stable change was occurring over time (as cited in Wardhaugh 2010, p. 206). Becker (2009, p. 646) also noticed that [r]-use has increased to the extent that even the lowest social class now produces this variable on an average of 36% of occurrences in casual speech. Becker (2009) discovered that "age emerges as a strong predictor of *r*-pronunciation," especially in the case of lower-middle-class speakers (p. 648). Becker (2009) found that speakers in the 30-39 year age-interval increased their [r]-use when compared to Labov's investigation of speakers' use of the same variable among members in the same age and class groups. Labov (1972, p. 137) originally found that this age group produced post-vocalic [r] 15 percent of the time during careful speech, whereas Becker (2009, p. 648) discovered that, after 40 years of progression, their age equivalents now produce the prestigious variant 26 percent of the time in interview speech. In other words, group members of the same age demographic have not only increased their overall frequency of use of post-vocalic [r] after forty years of real-time, but they have also adopted the prestige form associated with this variable to the extent that it now occurs with greater frequency in less formal contexts.

Change in Progress: The Age Factor

For more than 50 years now, sociolinguistic research has attempted to predict language change in progress (Becker, 2009; Chambers, 2002; Labov, 1972; Mather, 2010; Patrick, 2002; Trudgill, 1974; Wardhaugh, 2010). Most of the studies presented thus far have demonstrated that variations in linguistic behavior can mark stable differences in class, gender, and contextual styles in a variety of speech communities, which can lead to change. While this may be the case, Chambers (2002) argues that, although factors such as sex and class differences can indicate

fluctuations in language use and change, "when [language variation] marks change, the primary social correlate is age" (p. 355). For Chambers (2002, p. 355), whenever some minor variant occurs with increasing frequency from the oldest to the youngest generations, the change reveals itself in a prototypical pattern, commonly known as the S-curve. Whether or not we can identify a prototypical pattern of change in progress, research since Labov's 1966 seminal work has repeatedly shown that age is an important factor that can predict change in progress, even among members of the same social class.

In his analysis of speech variation by members of every working class group, Trudgill (1974) found that "age group differences will be more important than class differentiation," when attempting to identify language variation and change (p. 110). For example, data gathered by Trudgill (1974) revealed that older Norwich speakers retained the monophthongal pronunciation of his (a) variable (e.g. the vowel in *bad*, *bat*, and *carry*), whereas the same variable was undergoing a process of diphthongization by younger Norwich speakers. Trudgill's (1974) investigation is an example of linguistic change in progress in the sense that the newer form was marked by increasing frequency as the age of the participants decreased in apparent time. Trudgill (1974) concludes by arguing that the increase in diphthongal pronunciations by younger age-groups is evidence that a linguistic change is in progress.

Labov (1972) likewise found age to be an essential factor in identifying change in progress due to overt prestige values. Labov's (1972) use of the term prestige allows him to make the claim that newer forms have certain social significance attributed to them from either above or below the level of conscious awareness. This is because an interestingly complex relationship between age levels and class differences emerged when middle-aged members of the lower-middle-class were found to hypercorrect their formal speech patterns to approximate the

behaviors of younger, upper-middle-class speakers, who are believed to use forms that are newly prestigious (Labov, 1972). In short, though older members of the lower classes "tend to hold to the earlier norms…; it is only among the middle-aged groups that the new prestige norms were adopted" (Labov, 1972, p. 292). For Labov (1972), this is evidence that there exists a social mechanism that accelerates the introduction of some prestige linguistic feature, since the incoming variable can now function as an index of age level and group membership.

It may be useful at this point to explain what Labov defines as change from above or below the level of conscious awareness. When social pressures are imposed on language use, variations as well as changes to the language are presumed to be either conscious or unconscious – i.e. either above or below the level of conscious awareness, respectively (Labov, 2006, p. 203). The argument is that when speakers change their linguistic behavior without being aware of it, they do so unconsciously (from below), and vice versa for conscious changes (from above) (Labov, 2006, p. 204). In general, Labov's (1972) definition of change from above and below may be viewed as a pun. On the one hand, individuals have been seen to change their linguistic behavior due to the amount of conscious attention paid to speech; it is believed that vernacular is more likely to emerge below the level of conscious awareness when little to no attention is paid to speech, whereas the same individuals will change their speech to approximate more standard, prestigious forms above the level of conscious awareness when more attention is paid to speech.

The essence of the pun comes from an interpretation of the mechanism of language change in accordance with the values of prestige associated with particular linguistic variants; individuals consciously change their speech (from above) to approximate the speech of individuals they perceive to be positioned higher (or above them) in the social hierarchy. Labov (2006) illustrates this by describing two cases that frequently occur: (1) language features that

become socially stigmatized typically index uneducated or unrefined speech (i.e. they are associated with lower status groups); and (2) language features that are not used by the majority of the population are associated with prestige (which typically index higher status groups).

There are, however, instances when the change occurs from below, but is not necessarily associated with social status differentiations (Labov, 2006). This may also be called change from below, because successive generations gradually shift their linguistic behaviors without being consciously aware of it (Labov, 2006, p. 206). As Labov (2006) reflects, "perhaps it would have been better to call them 'change from without' and 'change from within,' since 'below' tends to be confused with the notion of lower class, but the terms are well established now" (p. 203). Nonetheless, when such a change does occur (from below), the shift is introduced by some particular group in society, which gradually generalizes the linguistic behaviors to other groups, so long as overt pressures from above do not superimpose some form of transformation (Labov, 2006, p. 206). Insofar as the situation is not swiftly corrected from above, "we would expect to see a steady progression along at least one social dimension as well as the dimension of apparent time" (Labov, 2006, p. 207). That is, at least one social dimension will elicit variations in language use, dependent upon which external factors are involved, and these variations can be observed by identifying age differentials in apparent time.

Also important is that when changes of this type occur from below, "there is no important distinction between stigmatized and prestige forms"; instead the form assumed by members of a particular group may unconsciously mark self-identification or group affiliation (Labov, 2006, p. 207). Labov (2006) demonstrates this in his recent analysis of the (æh) variable in the following class of words: *bad, bag, ask, pass, cash,* and *dance*. According to Labov (2006), the prestige value in this class of words is pronounced with a more open pronunciation (e.g. the vowel in a

word like *bat*, which has have the approximate phonetic quality of [æ:]), while the stigmatized variant exhibits a rise in height (e.g. the vowel in a word like *beer*, which have the approximate quality of [ɪᵊ]). Labov (2006) assigned six distinct values to this sociolinguistic variable. The variable (æh-1) is the highly stigmatized form and (æh-6) is the most prestigious form. The results showed that the speech of older generations in the upper-middle class tended to retain (æh-2) and (æh-3) during casual speech, whereas younger members of the same class produced (æh-4) during casual speech. In contrast, a reverse pattern was observed for lower classes, in which only the older generations retained the prestigious (æh-4) forms during casual speech. The results show that, rather than adopting the prestigious form, the lower social classes chose to adopt the stigmatized version to identify with members of similar social groups. Labov (2006, p. 240) refrains from providing a clear explanation for this phenomenon, save for a small footnote, suggesting that "although change from below gradually affects all classes, it appears to originate with some particular group; this group is most often a lower ranking social group, though not necessarily."

In another study which surveyed the relationship between age and language variation leading to change, Chambers (2002) explored the decline in [hw]-pronunciations among Canadians. In the case of the linguistic variable (wh), Chambers (2002) demonstrates, with data from a previous study, that the youngest age groups in central Canada have eliminated the use of the [hw] variant, whereas the oldest groups used [hw] more frequently (Chambers 2002). The middle groups demonstrated a trajectory of the change that steadily rises from the older to the younger generations, with an increase of [w] use occurring within about 10 percent of the population every decade for nearly fifty years in apparent time (Chambers, 2002, p. 361). Consequently, Chambers (2002) argues that change in progress occurs when there is evidence of

"incremental increases in the use of a particular variant in the speech of younger people" (p. 360). That is, younger age groups tend to exhibit higher frequencies in use of an innovative or incoming variant in apparent time, which can lead to changes in the language.

Because the factor of age is considered to be a primary social correlate linked to language variation and change in a speech community (Chambers, 2002), it is reasonable to suggest that the variability in electronically-mediated environments will correlate similarly with the age of the participants. The first important research question, then, is how does age correlate with the linguistic behaviors of EMC users when they use informal variants to represent features of speech and orality in the transmission of ETMs?

Style in Sociolinguistics

Thus far, I have presented a series of studies that examined how language variation and change are functions of external social factors with minimal reference to style. While age – among other social factors – is a primary social correlate that may indicate variation and change to a language (Chambers, 2002), stylistic variation also offers linguists the possibility to observe linguistic change in progress (Rickford & Eckert, 2001, p. 1). For Coupland (2007), "This means focusing on particular moments and contexts of speaking where people use social styles as resources for meaning making, [because] speakers project different social identities and create different social relationships through their style choices" (p. 3). That is, monolingual speakers have been seen to adjust their speech by choosing one or more varieties of the same language, and, in a number of cases, choosing one or more varieties does not solely depend on external social factors, but also on the context and on the speakers' sensitivity to the role of relationships expressed through language use (Coulmas, 2005, p. 57). The widely held conviction is that speech style and social context interrelate, because linguistic style is often variable when social

significance is linked to the speakers' perceptions of the particular context (e.g. the level of formality) and of group affiliations and personae associated with some linguistic form, such as differences in social dialects (Coulmas, 2005; Coupland, 2007; Rickford & Eckert, 2001). In this section, I will present a brief overview of modern theories regarding stylistic variation and its importance in observing variable linguistic behavior, especially in the case of formality.

One of the main critiques of sociolinguistic methodologies is that "the vernacular the linguist wishes to observe is unlikely to be produced in the relatively formal context in which speakers interact with interviewers who are strangers" (Rickford & Eckert, 2001, p. 3). Labov (1972) was fully aware of this social constraint on linguistic variation, as he called it the 'observer's paradox.' He sought to operationalize a wide range of interviewees' styles, primarily by using various devices to manipulate the topic and divert the participant's attention away from speech, thus allowing vernacular forms to emerge (Labov, 1972; Rickford & Eckert, 2001). These earlier models tended to view stylistic variation as a way for speakers to position their behavior with respect to social pressures (Rickford & Eckert, 2001).

More recent theories emphasize the need to treat stylistic variation as a dynamic presentation of the speaker's own personae in relation to the values they associate with the linguistic variables (Coupland, 2007; Rickford & Eckert, 2001). As Coulmas (2005) comments, choosing vernacular forms and colloquial styles are often a speakers' way to convey intimacy and solidarity with a group, and "the social system underlying these relationships influences [speakers'] speech" (p. 57). So, since modern sociolinguistic theories of language and style consider identities (both personal and social) and social meanings to be fundamental factors in stylistic variation, it will be useful to review both factors and their relation to linguistic variability.

According to Coulmas (2005, p. 171), one of the basic tenets of sociolinguistics is that a speaker's identity is displayed by their linguistic behavior. "It cannot be denied that language is experienced as a marker of identity, [since]… as we speak, we reveal who we are, where we grew up, our gender, our station in life, our age, and the group we want to belong to"; and we would be unable to do so if other people did not behave in a similar manner (Coulmas, 2005, p. 173). Both language and identity are often viewed as dynamic, continuously shifting and being renegotiated and co-constructed in response to the fluid and ever-changing contexts of our interactions (Coulmas, 2005; Llamas & Watt 2010). While this may be the case, there are instances in which the identity of the individual may be viewed as stable, such as the reoccurring use of particular social dialects that become associated with regional differences. This is because individuals can be seen to make use of the same linguistic features in a variety of contexts. In short, there are limits to the flexibility of individual and social identities.

Nevertheless, in the view explained by Coulmas (2005) and others (c.f. Bucholtz & Hall (2010), Coupland (2007), and Edwards (2009)), identity is believed to be neither static, nor does it emanate primarily from the individual psyche, but instead it is "a relational and a socio-cultural phenomenon that emerges and circulates in local discourse contexts of interaction" (Bucholtz & Hall, 2010, p. 18). The elements of individual identity are, therefore, typically seen as a combination of personality traits, characteristics and dispositions embedded in and reflected by social practices (Edwards, 2009), in which discourse practices play a significant role (De Fina, Schiffrin & Bamberg, 2006).

In sociolinguistics, particular language varieties are understood as markers of group identity (Chambers, 2002; Edwards, 2009). That is, personal characteristics and identities are socially constructed via socialization within the groups to which the individual belongs

(Edwards, 2009, p. 16; Squires, 2010). Bucholtz and Hall (2010) incorporate this view into a similarly broad definition of identity, stating that "identity is the social positioning of the self and the other" (p. 18). What is implied here is that markers that delineate group membership fall in line with the assumption that humans construct social identities through values and norms established and evaluated by the various groups to which they belong.

Delineating group membership by virtue of shared norms has been at the forefront of sociolinguistic research for nearly half a century (Chambers, 2002; Coupland, 2007). The appeal to factoring identity by interpreting trends and patterns of variable linguistic behavior dates back to Labov's 1963 study of phonological patterns on Martha's Vineyard (Coupland, 2007; Labov, 1972; Llamas & Watt 2010; Wardhaugh, 2010). While Labov (1972) refrained from an in-depth analysis of the effect of linguistic identity on language form, he nevertheless demonstrated how speakers' linguistic variability is a reflection of her or his identities. On Martha's Vineyard, group evaluations of a particular linguistic form were interpreted by Labov (1972) to function as a form of resistance to social pressures by virtue of marking solidarity with members of the island's distinctive culture. For instance, Labov (1972) observed how fisherman exhibited greater use of the centralized pronunciation of /aw/ and /ay/ to express resistance to mainlanders. To explain the phenomenon he was observing, Labov (1972), argued that when an individual orients her or his speech to these particular unifying subphonemic features while residing in the island, "he is unconsciously establishing the fact that he belongs to the island: that he is one of the natives to whom the island really belongs" (p. 36). That is, the more a person centralized the first part of the diphthong, the more s/he identified with those native to the island.

Since the time when Labov showed that style can be interpreted as a set of co-occurring variables associated with speakers' identities, attitudes, perceptions, and so on, alternative ways

of talking about identity prompted sociolinguists to question how variable linguistic forms acquire social meanings (Johnstone, 2010). Modern sociolinguistic theories focusing on speakers' identity, therefore, take into account the participant's perceptions of some linguistic style (Rickford & Eckert, 2001**)**. Coupland (2007, p. 9), for example, argues that analyses of how speech variants are normatively distributed across members of some social stratum abstracts away from speakers' perceptions of social processes in relation to the interactional contexts.

Style, Meaning, and Context: The Role of Formality

At this point, an operational definition of style is due. Style fundamentally refers to a particular way of doing something (Coupland 2007). In linguistics, style "refers to ways of speaking – how speakers use the resource of language variation to make meaning in social encounters" (Coupland, 2007, p. 32). Contemporary approaches to style in sociolinguistics understand style as a repertoire of linguistic features associated with personae and identities that are linked to particular contexts (Bucholtz & Hall, 2010; Johnstone, 2010). In this sense, style is perceived to be an additional mechanism that prompts people to exhibit linguistic variability for the sake of negotiating and renegotiating shifting identities according to the circumstances. Investigations of this sort are said date back to Coupland's, Hall's, Gumperz, and Hymes' conceptualizations of audience design and accommodation theory (Bucholtz & Hall, 2010; Coupland, 2007; Rickford & Eckert, 2001). The present study will not attempt to explore audience design and accommodation theory at work; instead I will build upon the theoretical underpinnings of stylistic variation – or as Coupland (2007) calls it, stylistics – to present interpretations of how speakers shift their style according to particular social contexts.

Once again, identifying a relationship between one's linguistic behaviors and varying levels of formality dates back to what Labov (1972) called stylistic variation. He came to

operationalize sociolinguistic style according to a hierarchical analysis of English style, which was greatly influenced by Joos (cited in Labov, 1972), and which endorses a linear scale of formality (Coupland, 2007, p. 40). The method used by Labov (1972; 2006) to observe different speech styles, specifically in relation to formality, was to provide interviewees with different contexts that can determine the amount of conscious attention the individual pays to his or her own speech. In other words, Labov (1972) views stylistic differentiation as the more conscious attention paid to speech; the more speakers were consciously aware of their speech, the more likely they were to use the formal variety and the reverse for when less conscious attention is paid to speech. Again, Labov (1972) attempted to isolate five different contextual styles that were believed to elicit different speech styles according to varying levels of formality (i.e. from the most casual to the most formal or careful types of speech). For Labov (2006), casual speech refers to the (typically vernacular) speech varieties used every day in informal situations, "where no attention is directed to language" (p. 64). Careful speech, in contrast, is the type of language that receives to full attention of the speaker; the context in which the most prestigious forms are believed to emerge.

Again, however, style in the Labovian sense traditionally referred to language use as it relates to intra-speaker variation (Labov, 1972). As a result, style has subsequently been treated quantitatively with the use sociolinguistic surveys in a similar way that variations associated with social class differentiations have been treated (Coupland, 2007, p. 9). "Formality or carefulness [was] assumed to be a matter of speakers modifying their speech in respect of those same features that define their place in a social hierarchy" (Coupland, 2007, p. 9). Thus, Coupland (2007) argues that the methods used to investigate the speech of individuals in interviews tend to

probe how an individual speaker produces some version of social style according to particular speaking situations.

Yet, linguists interested in how context shapes speakers' styles object to sociolinguistic surveys that limit the range of contexts in which style can be observed and analyzed (Coupland, 2007, p. 6; De Fina et al., 2006; Rickford & Eckert, 2001). Recent studies of style in variationist sociolinguistics insist that, in order to understand speaking and styling as sociolinguistic processes, it is necessary to look at the notion of social contexts (Coupland, 2007). In sociolinguistics, social context refers to the immediate social and physical settings that influence speakers' linguistic style. Sociolinguists maintain that social identities are indexed by speakers' language use in text and talk, and, thus, social contexts can point to certain "local (objective and subjective) norms that constrain linguistic style" (p. Coupland, 2007, p. 14). For example, the setting of a church may be linked to whispered talk or even silence (Coupland, 2007). Social context is, therefore, included in the culture in which the individual is positioned, since, at the macro-level, the culture can define the context for social interaction, which can then be specified according to the particular social situations and circumstances (Coupland, 2007, p. 14).

Much like identity, social contexts are multilayered and complex. "Incorporation of the context is in itself dynamic" (De Fina et al., 2006), because context is in part a phenomenon that is socially structured, and to which speakers must subscribe as they 'live it out' in their talk (Coupland, 2007, p 19). The argument made by Coupland (2007) is that the variationist study of social styles and dialects needs to orient to the styling of meaning in the context of social interaction.

So, in order to observe and analyze the styling of variable linguistic behavior in a range of social contexts, we must also be sensitive to the contextualization processes at play. The study

of social context typically focuses on the process of active, local meaning-making (Coupland, 2007; De Fina, et al., 2006), and in an undertaking of this sort, "it makes more sense to talk about processes of contextualization – sociolinguistic style *creating* context as well as responding to context" (Coupland, 2007, p. 17; italics in original). Coupland (2007) argues that it is important to keep in mind that different people may perceive some given social context differently, which may influence linguistic behavior due to the subjective and objective processes made relevant by the speakers and hearers.

Indeed the most common influence and product of how one styles their variable linguistic behavior is the degree of formality (Edwards 2009). As Edwards (2009) comments, there is certainly a difference, for instance, among the statements, "I ain't gonna do it," "I'm not going to do it," and "I prefer not to do it," yet a single speaker may be heard to utter all these statements in different settings or situational contexts (Edwards, 2009, p. 28). Although Edwards (2007) claims that "This sort of variation is so effortless that we usually do it without much thought, and most members of speech communities adapt all the time," (p. 24) it is important to recognize that variation becomes noticeable when it seems odd, inappropriate, etc. for the situation and context (Edwards, 2009).

To demonstrate how Labov's (1972) hierarchical analysis of style can be combined with Hymes' 1974 SPEAKING mnemonic, which describes the various components necessary to understanding how particular speech events are relevant to the speakers' and hearers' objectives, Coupland (1980) studied the significance of contextual constraints in relation to variable language use in what he considered to be more "natural data" (p. 2). Rather than selecting elicitation contexts to observe language variation in a sociolinguistic interview (as Labov (1972) had done), Coupland (1980) attempted to emphasize stylistic variation according to the

components of the context. For Coupland (1980), "combining these two basic approaches makes it possible to analyse statistically a speaker's linguistic behavior over a range of micro-contexts," such as setting, situation, objective, and so on (p. 3). Coupland (1980), therefore, argued that he was interested in analyzing language variation from a post-hoc classification of contextual components, which interpret stylistic variation dynamically instead of statically as an automatic correlate of the context.

The study took place in a travel agency in Cardiff, Wales in 1980, and the goal was to determine how the different contexts of telephone talk, combined with message content and participants' objectives, influences linguistic variability among travel agents (Coupland, 1980). Coupland (1980) isolated five sociolinguistic variables (h-dropping, (ou), (r) realization in intervocalic and word-initial positions separately, simplification of word-final consonant clusters, and intervocalic /t/) and found that certain participants "produced what appeared to be three distinct 'levels of standardness' across different speaking contexts in the course of [their] day-to-day work" (Coupland, 2007, p. 69). Audio-recorded talk of sales agents led Coupland (1980) to identify four levels of contextual components (casual, informal work-related, client, and telephone) associated with different frequencies of use of the linguistic variables. As in Labov (1972), casual speech was considered the context in which the least attention was paid to speech, most commonly associated with vernacular speech (Coupland, 1980). Informal work-related talk was identified as the context in which individuals spoke to their co-workers about work related topics in a somewhat casual manner (Coupland, 1980). Whenever the agents discussed work-related topics with the presence of a client addressee, these contexts were labeled as client talk, and, finally, telephone talk occurred when the agents exclusively spoke to unfamiliar colleagues about work-related topics over the phone (Coupland, 1980).

According to Coupland (2007), his identification of the context showed a stylistic "hierarchy of contexts based around something like 'formality'" (p. 70). The distribution of linguistic features across all four interactional contexts revealed that speakers produce most instances of the vernacular during casual speech, which diminished as they moved from informal to the two most formal contexts (client and telephone). Unsurprisingly, the most standard-like forms occurred with greater frequency in the two most formal contexts (Coupland, 2007).

These conceptualizations essentially expand upon Labov's earlier models of intra-speaker variation and bring to the fore the notion that a speakers' linguistic variation is also shaped by the norms associated with the interactional context in particular circumstances and situations (Bucholtz & Hall, 2010; Coupland, 2007; Edwards, 2009; Ervin-Tripp, 2001; Johnstone, 2010; Rickford & Eckert, 2001). Insofar as these socially conditioned norms reflect speakers' variable linguistic behavior in a variety of interactional contexts, it is possible to formulate the second research question central to this study**:** how do EMC users' perceptions of formality influence their decision to represent features of speech and orality with the use of informal variants in electronic text-based interactions? Answering these research questions will help inform the sociolinguistic inquiry of language variation and change with particular attention to the practices of EMC users. In the following chapter I will examine some of the very few research models that have attempted to bridge this gap.

Chapter 3: Speech Communities in EMC

Chapter Overview

To date, there exists a limited amount of research investigating how sociological factors and stylistic variation can mark group membership in virtual communities. Much sociolinguistic research has been concerned with the social and cultural context embedding particular instances of language use as it occurs in the spoken channel. Yet, variations in a language do not project possible changes to speech alone. What is affected is a language. It is, therefore, necessary to view language as a system. The users of this system have a tendency to establish markers of normative linguistic behavior originating from shared sets of knowledge, values and expectations about the system (Barnes, 2003; Danet & Herring, 2007; Labov, 1972). In this chapter, I will present some of the most salient features that have been researched, so as to form a basis for the argument that the language of EMC is highly variable, often shaped by social factors and social contexts.

Common Features of EMC

The language of EMC often mixes elements of writing and orality (Barnes, 2003). Some common features of EMC in English include rebus writing (e.g. *b4* for *before*), emoticons such as :) to represent objects or concepts (in this case a smile), colloquial variants to represent socially marked features of speech (e.g. *gonna* for *going to*), the use of exaggerated spelling and marks of punctuation (e.g. sooon!!!!!) to express paralanguage and prosodic cues for emphasis, and EMC-specific acronyms (e.g. *lol* for *laugh out loud*) (Anis, 2007; Herring 2001; Palfreyman & Al Khalil, 2007). Scholars, such as Barnes (2003), Crystal (2008) Herring (2001), and Danet & Herring (2007), among others, recognize that many of these linguistic practices have been adopted and continue to be used by EMC users as "a number of compensatory strategies to

replace social cues normally conveyed by other channels in face-to-face interaction" (Herring, 2001, p. 623). The argument made by Herring (2001) (cf. Barnes, 2003) implies that the norms and linguistic practices shared by EMC users' originate from features people acquire in non-virtual, off-line environments.

Social Evaluations of EMC Features and Users

Although some have argued that the linguistic practices of EMC users are more dependent on the physical constraints of the medium, Palfreyman and Al Khalil (2007) claim that "this downplays the social significance of this way of writing" (p. 46), specifically because sociolinguistic research has shown that many aspects of language (whether they are variations in speech or stylistic choices in writing) can be used as markers of in- and out-group identity (Chambers 2002; Coulmas 2005; Coupland 2007; Danet & Herring 2007; Edwards 2009; Herring 2001; Labov 1972; Labov 2006; Palfreyman & Al Khalil 2007; Trudgill 1974; Wardhaugh 2010). In Chapter 2, I outlined sociolinguistic research from Labov (1972; 2006) and Trudgill (1974), which found that the use or nonuse of some phonological feature can index an individual's position in social strata. Kiesling's (1998) study of the complex social meanings of the (ing) variable discovered that, rather than indexing uneducated social attributes, the use of the nonstandard [n] form can, in fact, index alternative social attributes, such as a casual, confrontational, or hardworking approach. And in Chamber's (2002) analysis of the (wh) variable, he found that the less frequent use of the [hw] was a marker of younger age groups membership within the speech community. Finally, with regards to formality and the social evaluation of linguistic variants, Coulmas (2005) (cf. Coupland, 2007) notes that vernacular, nonstandard forms are most often used in informal, intimate settings; "It is the colloquial style that conveys intimacy and solidarity, [whereas] a formal style [is] associated with power and

formality" (p. 127). In the present study, I propose that EMC users are drawing on similar off-line behaviors (specifically with respect to age and formality) when they use the vernacular or colloquial variants in ETMs more than other groups, and that the choices made by EMC users are influenced by social evaluations.

Though the concept that language variation in ETMs is influenced by social variation is sometimes overlooked in EMC research, Warschauer, Said, and Zohry (2007) examined interactional practices of 43 young Egyptian professionals between the ages of 24-36, who were communicating via email messages and online chats in both English and Arabic. They found that the "written form of romanized Egyptian Arabic is…widely used in informal communication" (Warschaer et al., 2007, p. 314). For example, the participants made widespread use of the number 2, 3, and 7 to represent phonemes that are "not easily rendered in the Roman alphabet" (Warschauer, et al., 2007, p. 312). In fact, in most formal online communication, many young professionals strongly emphasized Standard English in their survey responses (Warschauer, et al., 2007). The data support this claim, which show that 82.5% of the participants used Standard English in most of their professional writing, whereas in informal email correspondence, the majority of users code-switched and exhibited a higher frequency of use of the vernacular in both languages (Warschauer, et al., 2007). Classical Arabic, on the other hand, occurred much less frequently in the set of ETMs. The authors (Warschauer, et al., 2007) relate the social significance of using Standard English in online interactions to several factors, the most influential of which include social and economic pressures: educational and religious authorities frown upon the written use of Egyptian Arabic in areas such as religion, scholarship, and business (p. 312). In an attempt to relate their findings to Labovian interpretations of gradual language change, Warschauer et al. (2007) conclude that several socioeconomic changes

contribute “toward a shift from English diglossia in Egypt to multilingualism, with both English (from ‘above’) and Egyptian Arabic (from ‘below’) encroaching on the traditional dominance of Classical Arabic in written communication” (p. 315).

In another study, Ling (2005) was interested in a sociolinguistic analysis of the linguistic practices of 463 EMC users in a corpus of 867 text messages in Norwegian. His findings generated several implications about the users and their social evaluations of nonstandard linguistic variants (Ling, 2005). In terms of users, Ling (2005) discovered that females, teens and young adults were the most active users of the medium. In fact, younger teenage females appeared to have a broader register of linguistic variants when transmitting text messages than their male counterparts, despite their greater command of the standard; that is, young females showed a greater use of capitalization and message complexity than their male counterparts in the same sociodemographic categories. With specific reference to the greater use of abbreviations by younger females, Ling (2005), mentions that “the use of these forms of interaction also contribute to a sense of group…the use of various forms of abbreviation are seen as ways of identifying group membership” (p. 343). From these results, Ling (2005) suggests that younger females’ higher frequency of use of these features was viewed as a linguistic practice to index a more emotional and intimate side of mobile phone communication among young females.

So far, the studies presented in this paper, along with other sociolinguistic research, have claimed that linguistic variables function as markers of group membership in a given community (Chambers 2002; Labov 1972), which can serve as a reflection of the value system of the members within the same community (Trudgill, 1974). This claim has also been supported by research investigating the linguistic practices of EMC users (Anis, 2007; Danet & Herring 2007;

Herring, 2001; Palfreyman & Al Khalil, 2007), as shown in the studies above (Ling, 2005; Warschauer, et al., 2007). Much like offline settings, studies of interactive online communication indicate that the linguistic practices of users in EMC settings carry with them some social significance about language use and the attributes of the EMC users.

Social Factors, Style, and the Language of EMC

Some studies focus more on which type of media motivates variant forms of language use, or on how external social factors can be identifiable through content or through user names (Baron 2008; Herring 2001). Herring (2001), for example, argues that "Information about participants' *educational level* is given off largely unconsciously by their sophistication of language use, including adherence to prescriptive norms; similarly, *age* is often revealed through preoccupations and life experiences communicated in the message content" (italics in original p. 621). Whether or not these claims can be substantiated through quantitative and/or qualitative research, Herring (2001) argues that the recent phenomenon of human-to-human communication via systems of computer networks raises issues for traditional variationist methods, since reliable and, more importantly, clearly identifiable information about age, social class, gender, race, etc. are difficult to determine (Herring 2001). Nevertheless, there is evidence that users make use of variable linguistic practices to compensate textually for missing gestural and auditory cues (Herring 2001).

To demonstrate this latter point, Danet and Herring (2007) compiled 15 empirical studies of online communication in a variety of languages and EMC modes. The findings showed clear evidence that several informal writing conventions of EMC occur in multiple languages besides English. Though more synchronous modes of EMC tended to exhibit features that are more informal and "speech-like" than more asynchronous ones, several features are shared across

various modes (e.g. emoticons, abbreviations, etc.) (Danet and Herring 2007; Herring 2001). In fact, much like general findings regarding variable linguistic behavior in offline settings, age and degrees of formality tend to influence the use of these and other linguistic variants (Danet and Herring 2007; Herring 2001; Page 2012; Palfreyman and Al Khalil 2007).

Similar to Warschauer et al.'s (2007) study, Palfreyman and Al Khalil (2007) found that Gulf Arabic teenagers made frequent use of both Arabic and Roman alphabet writing systems to employ an array of conventions to link written symbols with sounds of spoken Arabic. For example, "< ز > is replaced by /z/ on the basis of pronunciation, rather than <j>, which looks similar," and "the numeral <5> is also used as an alternate to <′7> to represent the sound /x/ [which] appears to derive from the fact that the Arabic word for 'five,' /xasma/, begins with the same sound" (Palfreyman & Al Khalil, 2007, p. 54). Quantitative analysis of an authentic corpus of Instant Messenger (IM) text messages showed that: (a) the register is generally informal; and (b) turns are typically short (often no more than four words), similar to informal colloquial speech (Palfreyman & Al Khalil, 2007).

The authors then gathered qualitative insights from a follow-up attitudinal survey. In response to the attitudinal survey, one participant commented that the language found in EMC is a "kind of code, we feel that only ppl of our age could understand such symbols and such way of typing…i guess it's a funky language for teenzz to use" (Palfreyman & Al Khalil, 2007, p. 59). Similar responses from the survey led Palfreyman and Al Khalil (2007) to argue that such aspects of language act as markers of in- and out-group identity, since the desire to represent spoken features textually "is a result of social pressure to break conventional spelling rules and comply with IRC's nonconformist… image" (p. 46). This is because the use of vernacular,

according to Palfreyman and Al Khalil (2007), is believed to be a symbolic resource for informal communication among young Gulf Arabs.

Referring back to Ling's (2005) study, age and gender seemed to factor into: (a) which types of people are more frequent users the medium; and (b) which users are more likely to use alternative spellings and orthographic conventions to represent speech in writing. With respect to which age groups exhibit a higher frequency of use of the medium, Ling (2005), found that 85% of teens (the two youngest age cohorts were divided into 13-15 and 16-19 year-olds) and young adults (those in the 20-24 age range) reported sending text messages daily; they "are more adroit users" of the medium (p. 348). To analyze which groups are more likely to use linguistic practices that are believed to approximate speech, Ling (2005) examined the following variables: the lack of structural complexity in terms of length of the message; the use of abbreviations; punctuation; and capitalization. Except for the 20-24 age group, which appeared to be the most prolific users of punctuation – keep in mind that the frequent use of punctuation marks were attempts to include emphasis in the messages, such as ellipses for dramatic pause (…), exclamation marks to indicate excitement or surprise, and multiple question marks to indicate "advanced confusion" (Ling, 2005, p. 343) – and capitalization, the two teenage groups outperformed all other groups in the remaining variables laid out by Ling (2005), and the frequency of use declined rapidly with an increase in age.

Returning to the variability of EMC language in English, Tagliamonte and Denis (2008) attempted to show how the linguistic practices of teenagers participating in IM conversations can demonstrate a gradual change in progress. The authors identified 16 of the most frequent linguistic forms (or variables) they believe to be distinctive to EMC (e.g. *haha* for laughter, *omg* for *oh my god, btw* for *by the way,* etc.). From these, Tagliamonte and Denis (2008) isolated

three highly frequent forms (*lol, haha,* and *hehe*), and noticed that frequency of use of the acronym *lol* (for *laugh out loud*) and *hehe* was increasing among younger teenagers in the 15-16 age range, whereas older teenagers in the 19-20 age range retained a clear preference for *haha*. The authors also investigated how forms of intensification (e.g. *very, so*, and *really*) are undergoing a linguistic change in IM (Tagliamonte & Denis, 2008). When they compared their EMC data with a corpus of spoken English taken from two teams of teenagers in a three-year research project led by the first author (Tagliamonte) between 2004 and 2006, Tagliamonte and Denis (2008) found that *very* is frequent among older teenagers, while *really* is increasing among younger teenagers, and *so* appeared to be the newest form, "on the horizon" of the linguistic change (p. 17). This led them to conclude that "our results corroborate earlier CMC research in demonstrating that language use in IM is part of a much broader trend toward more informal language generally," because formal variants are seemingly undesirable in teenage text-based conversations (Tagliamonte & Denis 2008, p. 27).

Though not all aspects of personal identity are visible in online interactions, research shows that an individual's identity and her or his membership with one or more groups can influence language use in EMC. There is evidence that the language practices of EMC users can be seen to index identities that are socially conditioned by features associated with one or more social groups. Given that such variation reflects the influence of social factors – such as situational context and participant sociodemographic factors – on the linguistic choices of EMC users, sociolinguistic inquiries exploring language variability among EMC users need to recognize that offline and online behaviors can overlap (Danet and Herring 2007; Herring 2001; Page 2012). These insights will help in the development of a hypothesis for the current study. However, prior to offering any proposed notions of EMC users' language use, a more in-depth

description of the features that are believed to form group memberships in EMC environments is essential. This will be the focus of the following chapter.

Chapter 4: Representing Speech and Orality in EMC

Chapter Overview

In order to undertake a sociolinguistic investigation of how EMC users' linguistic behaviors vary according to age differentials and two levels of formality, it is necessary to identify a set of sociolinguistic variables. The focus of this chapter is to present some common trends and practices shared by EMC users whenever they produce graphic representations of a variety of linguistic features characteristic of speech and orality. First, I will briefly identify certain characteristics that are said to distinguish speech from writing. I will then discuss the correspondence between written and spoken features in relation to traditional spellings. Thereafter, I will examine some of the linguistic processes that play a significant role in users' attempts to display features of both written and spoken language in EMC, as well as ones that are nonverbal and considered "distinctively digital" (Danet & Herring 2007, p. 12). Finally, I will discuss some relevant terminology as it relates to the English writing system, outlined by Sampson (1985) and Coulmas (1989). The terminology will aid in the classification of linguistic features (defined in chapter 5), which will become the three dependent sociolinguistic variables for this study.

Speech and Writing

A common approach to understanding language use in EMC is to determine which side of the "speaking/writing divide" the linguistic features correspond to most (Crystal, 2001, p. 28). According to Crystal (2001), the most fundamental factors that differentiate speech from writing indicate that speech is typically face-to-face, time-bound, spontaneous, immediately revisable, loosely structured, oral/aural and socially interactive; whereas writing is typically space-bound, elaborately structured, visual and repeatedly revisable (Crystal, 2001, pp. 25-28). Adding to the

list, Baron (2008, p. 47) argues that there exist differences in structural properties (e.g. dialogue vs. monologue, ephemeral vs. durable, vagueness vs. precision, etc.), sentence characteristics (e.g. shorter vs. longer units of expression, simple vs. complex structure, etc.), and vocabulary characteristics (e.g. narrower vs. wider lexical choices, high vs. low use of slang, contractions, first person pronouns, etc.) from which speech can be compared to writing respectively. In all their attempts, scholars are usually left with the impression that language use in EMC "mixes elements of literacy and orality together" (Barnes, 2003, p. 92), because it is a "trans-modal phenomenon with features of both spoken and written cultures" (Ling, 2005, p. 347), and, as a result, it is "identical to neither speech nor writing, but selectively and adaptively displays properties of both" (Crystal, 2001, p. 47). Overall, it appears that searching for an appropriate method to divide language use in EMC according to properties of speech or writing is debatable. In the current study, I will focus more on the wide range of informal variants used by EMC participants to represent aspects of speech and orality in writing. First, a few words of caution regarding the linguistic "novelty" of such variants is necessary.

Linguistic Novelty?

While the varieties of informal language use reflect a great number of resources adopted by many EMC users, not all of the forms documented by EMC scholarship can be considered linguistically novel particular to the electronic written channel (Crystal, 2008, p. 37). How, then, do all these informal variants creep into language use in EMC? With respect to abbreviations, acronyms, and blends (explained below), the most straightforward answer is that EMC users are employing word formation strategies that they have likely acquired from other sources (Baron, 2008; Barnes, 2003; Danet & Herring, 2007). When people refer to the frequent use of abbreviations and acronyms that presumably belong exclusively to EMC (Crystal, 2001; Herring,

2001), researchers tend to overlook the fact that these linguistic devices have been used in other written contexts, including personal note-taking practices and telegraph messages (Baron, 2008; Squires, 2010) (e.g. *re* for *regarding*, *ant* for *antenna*, *rtty* for *radio-teletype* etc.). Certainly, there are a few innovative acronyms that appeared with EMC technologies (e.g. *brb* for *be right back*); however, when scholars research the language of EMC, they neglect to recognize that these forms of language use have existed for centuries or more (Crystal, 2008).

A similar argument could be made for the most commonly referenced informal colloquial variants, (e.g. *gonna*, *wanna*, *kinda*, etc.), in that nonfiction English authors have used these features for well over a century, which may have helped in conventionalizing the spellings. Thus, it is an inaccurate assumption that the wide range of linguistic features are novel conventions; instead, they are forms of language use which EMC users have adopted and expanded in their electronic interactions. Nevertheless, in order to understand how users make use of alternative spellings to produce informal variants for the sake of representing speech and orality in EMC, it is important to understand the relations between sounds and the typical spellings of words.

Traditional Spellings

There is a complex relation between the ways in which sounds are represented by transcription and the typical spelling of words in English (Huddleston & Pullum, 2002, p. 17). According to Huddleston and Pullum (2002), "When we match up written and spoken forms we find that in the simplest cases one letter corresponds to one sound, or **phoneme**: *in* /ɪn/, *cat* /kæt/, *help* /help/, *stand* /stænd/, and so on" (bold in original p. 17). For clarity, phonemes are contrastive, abstract linguistic units of sound, which distinguish words, such as the difference between /p/ and /b/, which distinguish the words *pit* and *bit* respectively (Cruttenden, 2001, p. 42). While one letter can, at times, correspond to one sound (or phoneme), the matching up of

written and spoken forms is often more complex, especially when there is a less direct relation between the letters and their correspondence to phonemes (Huddleston & Pullum, 2002, p. 17). For example, in the noun *breath*, the two letter sequence *ea* corresponds to /e/ and *th* to /θ/; in *night*, the three letter sequence *igh* corresponds to /aɪ/ (a diphthong, classified as a single phoneme in English); and in *through*, the four letter sequence *ough* corresponds to the single phoneme /u/ (Huddleston & Pullum, 2002, p. 17). Huddleston & Pullum (2002) use the term "symbol" as a single "unit of writing that corresponds to a phoneme, and we will refer to those symbols consisting of more than one letter as **composite symbols**" (bold in original, p. 17). Thus, in the examples above, the sequence *ough* in *through* is considered a composite symbol, whereas the letter <e> in *help* is simple (Huddleston & Pullum, 2002). I will reference some of the concepts in the analysis of the features identified in the later sections.

First, it is important to keep in mind that "speech is a continuous stream of sound without a clear division into units, but it can be analysed into meaningful elements which recur and combine according to rules" (Biber, Johannson, Leech, Conrad, & Finegan, 1999, p. 50). In writing, on the other hand, the analysis of meaningful elements is typically expressed through the division of orthographic symbols into words, as well as sentences (Biber et al., 1999), with the use of spacing, capitalization, marks of punctuation, and so on (Crystal, 2005). In EMC, however, the division of symbols into words does not always conform to the prescribed rules of spelling and word formation. As a result, it may be helpful to understand some basic forms of words as well as their characteristics and classes.

Words and Word Classes

In linguistics, the term morphology is used when discussing the forms of words (Crystal, 1997, p. 249), as well as the rules of word formation (Huddleston & Pullum, 2002, p.4). At the level of morphology, words are composed of morphemes and have the potential to be complex units of a language. For clarity, morphemes are considered the smallest units that carry meaning in a language (Biber et al., 1999; Huddleston & Pullum, 2002; Shockey, 2003). For example, the word *pins* represents the morphemes {pin} + {PLURAL} (Delahunty & Garvey, 2010). Thus, morphemes may be an entire word (as in *pin*) or they may be an element of a word (such as the –*s* inflection that represents plural). The morpheme *pin* is called a free morpheme because it can stand alone as a word, whereas the plural inflection –*s* (among other inflectional and derivational affixes) is called a bound morpheme, because it cannot (Huddleston & Pullum, 2002).

For ordinary language users, however, words are typically seen as "basic elements of a language" (Biber et al., 1999, p. 51) or as "individual items of vocabulary" (Huddleston & Pullum, 2002, p. 1623). Yet, as Biber et al., (1999, p. 51) mention, the definition of word is not as straightforward as it may seem:

> The independence of words is shown phonologically by the fact that they may be preceded and followed by pauses; orthographically by their separation by means of spaces or punctuation marks; syntactically by the fact that they may be used to stand alone as a single utterance; and semantically by the possibility of assigning to them one or more dictionary meanings. (p. 51)

How exactly these types of criteria can be applied differs according to the word.

According to Biber et al. (1999), in the examination of words, it is possible to broadly group words into three classes – lexical words, function words, and inserts – depending on their grammatical behavior and primary functions (Biber et al., 1999). Generally speaking, lexical words carry the most meaning in a text, whereas function words serve two major functions: (1)

they indicate how lexical words or larger units of language should be interpreted; and (2) they indicate "the relationships between lexical words or larger units" of the language (Biber et al., 1999, 55). Typically function words consist of a single morpheme and, as a result, can rarely be separated into smaller elements that carry meaning, whereas lexical words often, though not entirely, have a complex structure, resulting from word formation processes (Biber et al., 1999, p. 57). Another important distinction between lexical words and function words is that the former are usually stressed in speech, whereas the latter are not (Biber et al., 1999).

As for inserts, Biber et al. (1999) suggest that they are a relatively new category of word, which are free from syntactic structure (they can be inserted somewhat freely in a text), and which are marked off by pauses, intonation, or, in writing, by marks of punctuation. The examples provided by Biber et al. (1999) include *hm, uhhuh, ugh,* and *yeah* (p. 56). According to EMC research, representing graphically the features which Biber et al. (1999) call inserts is a common practice for EMC users (Crystal, 2001; Herring, 2001; Tagliamonte & Denis, 2008). Although it can be debated whether inserts should be regarded as words at all, there is little doubt that they "play an important role in communication [since]…They characteristically carry emotional and interactional meanings and are especially frequent in spoken texts" (Biber et al., 1999, p. 56). As a result, in order to adequately describe spoken language – and, thus, the representation of spoken language – attention needs to be paid to inserts.

Another important characteristic distinction that needs to be taken into account includes orthographic words. In general, orthographic words are written word forms that correspond to individual units of speech, but which are separated by spaces (and occasionally by punctuation marks) in written texts (Biber et al., 1999). With regards to transcribed speech in corpus material, it is possible to "find orthographic words that must be analysed as representing a sequence of two

grammatical words: *that's*, *wasn't*, etc." (Biber et al., 1999, p. 54). In EMC, it is not an uncommon practice to represent sequences of words as one orthographic unit, as evidenced by the examples *gonna*, *wanna*, etc. In fact, there are sequences in English, in which distinct orthographic words can "function together as single grammatical words during speech, e.g. *sort of*, in the conversation text" (Biber et al., 1999, p. 54), which are sometimes represented by EMC users as a single unit of speech (e.g. *kinda, sorta*).

In the sections that follow, only some of the most salient features documented in EMC and CMC research will be addressed, with reference to some of the linguistic processes that influence their nonstandard or informal forms. Thereafter, I will address other, typically informal word formation processes, by looking at the following categories: (1) abbreviations; (2) acronyms; (3) blends; and (4) rebus writing. Note that the categories under which these features fall will not form the basis of the sociolinguistic variables; instead, they will serve as an organizational method for interpreting how the forms of some commonly remarked features of EMC language have developed via several linguistic processes. These features will then be subsumed into three sociolinguistic variables on the basis of how the English writing system can be distinguished graphically.

Informal Emphatic Variants

There is little argument that EMC users draw upon a number of resources to provide linguistic information in the domains of prosody (i.e. pitch, stress rhythm, tempo, and loudness) and paralanguage (i.e. tone of voice, laughter, facial expressions, etc.), in order to "convey social and affective meanings" (see examples below) (Danet & Herring, 2007, p. 12). Crystal (1997) mentions that such features of suprasegmental phonology and phonetics are a means for speakers to signal "attitude, social role, or some other language-specific meaning" (p. 277). Indeed EMC

users are capable of representing auditory (and visual) information considered absent in online environments with the use of repeated letters (*sooooo, oooops, aaaaahhhhhhh*), repeated punctuation marks (e.g. *No!!!!, well…, what???*, etc.), emoticons (e.g. ☺, B-), :-/, etc.), and other emphatic conventions, such as the following (taken from Crystal, 2001, p. 35; and Herring, 2001, p. 617):

1. capitalization for shouting: I WILL NOT
2. orthographic representations for laughter: hahahahahahaha
3. letter spacing for "loud and clear": H O W C O M E
4. emphasis by asterisks: my *true* friend

The examples above are evidence that EMC users have adopted a wide range of devices to represent the dialogic spontaneity of spoken as well as nonverbal language in ETMs (see Chapter 3 for additional examples of common features in EMC) (Barnes, 2003; Fouser et al., 2000). According the Fouser et al. (2000), many of these features in EMC "represent attempts to convey orality…in online messages, which remain in written mode" (p. 52). For Fouser et al. (2000), the use of orthographic devices to represent gestures and other extralinguistic devices are not always equally distributed across speech acts (for clarity, Crystal (1997) states that speech act is a term that refers to how utterances can be analyzed "in relation to the behaviour of speaker and hearer in interpersonal communication" (p. 358)). So, for the purposes of the current study, the term orality is used to include features of the spoken or oral language that may not be considered to be directly tied speech or speech acts.

Representing Speech

Many of the conventions in the examples above and in chapter 3 are typically viewed as a mixing of elements of orality and literacy in EMC (Barnes, 2003); however, there is also evidence of alternative spellings and orthographies used specifically to represent speech and pronunciation in EMC (Anis, 2007; Crystal, 2001; Herring, 2001; Tagliamonte & Denis, 2008).

As Crystal notes, (2005), "Stylistic variation can be illustrated by the way authors adapt the spelling system to reflect or suggest the pronunciation of non-standard speech," as in the use of *sez* for *says*: the two forms are said to represent the same pronunciation, yet the former "conveys the impression of a non-standard accent" (p. 116). The number of features presented in this chapter is by no means a complete list of the informal variants that exist in EMC. Instead, only the most salient features will be addressed.

Scholars have attempted to advance terminology and provide operational definitions of the types of informal variants found in EMC considered unacceptable or inappropriate in formal written texts (Anis, 2007; Baron, 2008; Crystal, 2001; Crystal, 2008; Herring, 2001; Palfreyman & Al Khalil, 2007; Tagliamonte & Denis, 2008). Anis (2007), for example, uses the term "neography" to describe language forms that are typically unacceptable in formal writing, including features such as rebus writing, reduplication of letters, repeated punctuation marks, and so on. Further, Anis (2007) states that "a user can encode a phoneme or a sequence of phonemes; an alternative alphabetic system simplifies relations between graphemes and phonemes, and thus can be qualified as phonetics oriented" (p. 97). Hence, Anis (2007) uses the term phonetic spellings to refer to the orthographic transcriptions of standard pronunciations (e.g. the use of *nite* instead of *night*, or *guyz* instead of *guys*) or "socially marked variants ('luv' for 'love,' 'wanna' for 'want to')" (p. 97). Palfreyman and Al Khalil (2007) prefer the notion of phonological simulation when discussing the "representation of spoken features in online text, for example, the written use of English contractions such as 'gonna' and 'wanna'" (p. 46). Similar to Anis (2007), Palfreyman and Al Khalil (2007) maintain that such features display the importance of social norms in spelling and orthography due to a conventional desire to make written language seem as "speech-like" as possible (p. 46). Fouser et al. (2000) simply regard the

linguistic forms as deviations from standard spellings and orthographies, whereas others consider EMC to be a unique register (Squires, 2010), which exhibits "a fusion of the full range of variants…formal, informal, and highly vernacular" (Tagliamonte & Denis, 2008, p. 3). For the purposes of the present study, the term informal variants will suffice when discussing the overall tendency to produce written forms that are considered inappropriate or unacceptable for standard formal writing.

Corpus-based research on spoken conversation also includes notions of how particular informal, vernacular features can be treated once transcribed. Certainly corpus material relies on transcriptions of spoken conversation, or "speech rendered in written form" (Biberet al., 1999, p.1038). Nevertheless, in their survey of the main vernacular features of spoken conversations, Biber et al., (1999) describe several phenomena, many of which they categorize according to morphophonemic and morphological features. Regarding morphological features, Biber et al. (1999) refer to the use of the reduced forms: *ain't* (which corresponds to the standard informal contractions *isn't*, *aren't*, *hasn't*, *haven't* and *'m not*) and its range of pronunciations /ɛnt/, /ɪnt/ or /ɛn/; *innit* (the reduced form of *isn't it* and/or a further reduction of *ain't it*) pronounced /ɪnɪt/; *y'all* (the contracted reduced form of *you all*) pronounced /yɔl/; and the past tense forms of verbs considered irregular, such as the use of *drug* (pronounced /drʌg/) instead of *dragged*. As for the morphophonemic variants, Biber et al. (1999) discuss the colloquial use of the phonologically reduced pronouns *me* (/mɪ/) instead of *my*, *'em* (/əm/) instead of *them*, *ya* (/jə/) instead of you, as well as the use of *-in* (/ɪn/) instead of *–ing* found in gerunds and verb suffixes. In each case, Biber et al. (1999) note that these features show up "at the level of spelling, reflecting a different pronunciation from the standard one" (p. 1122).

Biber et al. (1999) comment that the terms nonstandard and vernacular can sometimes be "misleading in suggesting a clear cut dichotomy between two varieties of language: one which matches up to the 'standard' one and one which does not" (p. 1121). Yet, the term vernacular refers to a variety of a language that is found in colloquial speech, which is considered to lack overt prestige and to be inappropriate for formal written communication (Biber et al., 1999). In general, spoken conversation contains a wide range of vernacular, informal speech, which, according to Biber et al. (1999), "is often marked by effort-reducing features such as elision and assimilation" (p. 1048). Although many effort-reducing features are not (entirely) directly visible via traditional orthography, it is possible to observe reduced forms at the level of spelling (e.g. *them* /ðəm/ → *em* /əm/). Cruttenden (2001, p. 236) similarly notes that forms exhibiting simplification and reduction are typical of colloquial speech, whereas standard spellings influence formal speech to retain the fuller, more standard forms.

For the purposes of this study, I will focus on the informal varieties of language found in colloquial speech, especially since these features "can be highly prized because of their role in establishing and maintaining solidarity among the speakers in selected groups, and in bringing vigour and colour into speech style" (Biber et al., 1999, p. 1121). In the following analysis of informal variants, I propose that similar, colloquial conventions are used by EMC users whenever they choose to represent features of spoken pronunciation, perhaps for the sake of establishing and maintaining solidarity with members of the same group.

Informal Colloquial Variants

This section presents a sample of colloquial variants observed in English EMC research. In the current study, the term colloquial variant will be used when referring to the nonstandard linguistic forms of words that are shown to represent speech and pronunciation specifically. The

term colloquial describes the type of language speakers' use in a spoken conversation with familiar interlocutors, and, thus, is marked as informal. Although Tagliamonte and Denis (2008) argue that some highly informal and "reduced variants, including *gonna* can be considered vernacular" (p. 20), I will refrain from using the term vernacular variants for the current linguistic analysis of informal variants, since vernacular often refers to the dialect of a particular people and place (Biber et al., 1999). In general, it is difficult to link highly colloquial variants, such as *gonna*, to a particular region or dialect of English, due to the lack of alphabetic characters that can represent distinct phonetic features of a specific dialect accurately and completely. Nonetheless, according to EMC research (Anis, 2007; Crystal, 2001; Crystal, 2008; Palfreyman & Al Khalil, 2007), EMC users sometimes represent standard pronunciations (e.g. *nite* for *night*, or *guyz* for *guys*) or variants that are socially marked (e.g. *gonna* for *going to*, or *kinda* for *kind of*) to form words that have not become acceptable for formal, Standard English writing. Let us begin with an analysis of the latter type.

Colloquial Variants Influenced by Assimilation

In general, "the more common an item is, the more likely it is to reduce, given that it contains elements which are reduction-prone" (Shockey, 2003, p. 14). As a result, in all known languages, the articulation of "difficult articulatory sequences" is often simplified in connected, typically informal speech (Collins & Mees, 2003, p. 203). The term 'phonetic conditioning' is used to describe the ways in which the phonetic context influences realizations of phonemes due to pressures from adjacent (or near-adjacent) sound segments (Collins & Mees, 2003, p.203). The phenomenon commonly referred to as assimilation (Collins & Mees, 2003; Cruttenden, 2001; Crystal, 1997; Schreier, 2005) is an adjustment characteristic of connected speech (Celce-Murcia, Brinton, & Goodwin, 2010, p.163), which is one of the main types of phonetic

conditioning – elision is another type which will also be considered here (Collins & Mees, 2003, p. 203).

Broadly speaking, assimilation refers to the influence one sound segment has on the articulation of another (usually adjacent) sound segment, so that the sounds become more similar, or identical,(Celce-Murcia et al., 2010; Collins & Mees, 2003; Cruttenden, 2001; Crystal, 1997; Schreier, 2005). For example, when a word like *encounter* is accepted as an abstracted linguistic unit, its isolated, or ideal (typically artificial), realization [ɪnˈkaʊntər] differs from those when, in informal connected speech, the pressures from its sound environments or from rhythmic patterns are likely to produce the phonetic realization [ɪŋˈkaʊntər], i.e. the assimilated form (Celce-Murcia et al., 2010, p. 168; Cruttenden, 2001). The ideal form is used here to refer to the linguistic variant that is produced in the most careful styles of speech, which, "often bears resemblance to the spelling representation" (Collins & Mees, 2003, p. 204). In contrast, the form more typical of connected speech is referred to as the assimilated form. Since processes of assimilation can be analyzed according to several patterns, only the particular processes that pertain to the specific features under investigation will be addressed as they are introduced below.

Consider the following colloquial variants that may be regarded as influenced by assimilation: *hafta*; *gotcha* (also *gotchya*); *waddya* (also *whaddya*); and *watcha* (also *watchya, whatcha,* and *whatchya*), of which their respective standard written forms are *have to*, *got you*, *what do you*, and *what you*. In the case of *hafta*, the form appears to represent a type of regressive assimilation – i.e. features of a preceding, articulating sound segment anticipate, and are affected by, qualities of the following conditioning sound (Celce-Murcia et al., 2010; Collins & Mees, 2003; Cruttenden, 2001; Crystal, 1997). With some types of voicing assimilation, word-

final voiced fricative consonants [v,ð,z] may be realized as the corresponding voiceless fricative consonants [f,θ,s] in a sequence of sounds (Cruttenden, 2001, p. 283). So, in a phrase like, *I have to go*, the voiceless qualities of word-initial /t/ in *to* may condition word-final /v/ in *have* to be realized as the corresponding voiceless fricative [f], which has been transcribed with the letter <f>. In the formation of a word, then, an individual might prefer to spell *have to* as a single orthographic word, while also replacing the *ve* in *have* with <f> to represent the phonological processes of assimilation.

In the case of *waddya*, the form appears to be another type of regressive assimilation; however, in this case, the voicing of word-initial /d/ in *do* influences the voicing of word-final /t/ in *what*, which is then transcribed with the letters <dd>. The choice to duplicate the letter <d> may simply be due to the relation between spelling rules and the features of sound associated with them. In terms of spelling, common lexical word formation processes (whether by means of inflection or derivation) generally require that whenever a base ends in a simple consonant symbol, and is preceded by a simple vowel symbol, "the last letter [is] doubled before a vowel provided the base is stressed on the final syllable" (Huddleston & Pullum, 2002, p. 1675). Given that *what* meets these criteria, the spelling representing the informal pronunciation of the sequence *what do you* is influenced by ordinary spelling rules that determine (whether consciously or not) the doubling of the consonant <d> in *waddya*.

With respect to *watcha* and *gotcha*, both of these features appear to have undergone a process of coalescent assimilation, because they involve two adjacent sound segments that share a mutual two-way exchange of articulation features which fuse to create a third sound (Celce-Murcia et al., 2010; Collins & Mees, 2003; Cruttenden, 2001). In English this type of assimilation most frequently occurs during palatalization: i.e. when a syllable- or word-initial

palatal, such as /j/, is preceded either by a word-final alveolar consonant /s,z,t,d/ (Cruttenden, 2001, p. 286) or by the final alveolar consonant sequences /ts, dz/, the exchange of articulation features creates the palatal affricates [tʃ, dʒ] or the palatal fricatives [ʃ,ʒ] (Celce-Murcia et al., 2010, p. 171). Therefore, the two-way exchange of word- and syllable-initial palatal /j/ in *you* preceded by the word-final alveolar consonant /t/ in *what* and *got* influences the articulation of both features to coalesce, thereby resulting in the realized forms [ˈwɑtʃuː] and [ˈgɑtʃuː], respectively. The decision to represent this sound with the two-letter sequence *ch* is perhaps due to the sound associated with this sequence in several other words in English (*cheese* /tʃiz/; *church* /tʃərtʃ/; *chess* /tʃɛs/, and so on).

Colloquial Variants Influenced by Elision

Elision is a process of reduction that refers to the omission of some sound segment, which can lead to the loss of phonemes, syllables and, at times, words (Cruttenden, 2001; Crystal, 1997). Typically, when elision occurs at or near word boundaries, vowels and consonants are elided (Note: elision can also take place word-internally) (Cruttenden, 2001). Some of the most common forms of elision are presented here with reference to common colloquial variants in EMC.

In general, unstressed grammatical words, such as *of* and *and* are particularly prone to elision (Crystal, 1997). In the case of *of*, the phoneme /v/ tends to elide to in many varieties of informal speech. In the sequence *kind of*, for example, word-final /v/ in *of* may be elided and reduced to [ə], and the grammatical unit is sometimes orthographically transcribed as the colloquial variant *kinda*. A similar process of reduction can be found when *have* is unstressed and used as a "co-member of a sequence of auxiliaries" (Hudleston & Pullum, 2002, p. 105), for example, to form the perfect aspect in a sequence like *She could have gone*. In this case, word-

final /v/ in *have* may be elided and the sequence may sometimes be orthographically transcribed as *She coulda gone* (similar colloquial variants include *shoulda*, *woulda*, *mighta*, and so on). With regards to *and* (phonemically transcribed as /ænd/ in isolated form), the form may be elided to [ən] or even [n], and, as the example provided by Crstyal (1997) demonstrates, "the *a* and *d* are dropped in *boys 'n' girls*" (p. 133). Common spelling variants of *and* tend to represent the phonetically reduced forms with *an* or *n*, with or without an apostrophe.

Similarly, when common verbal forms, such as *want to* (as well as *got to* and *ought to*), form the word-final and word-initial sequence /tt/ in connected speech, the sequence is normally reduced to [t] (e.g. *We want to* /wiˈwɑnttu/ → [wiˈwɑntə], *I got to* /aɪˈgɒttu/ → [aɪ ˈgɒtə]) (Collins & Mees, 2003, p. 212). Specifically in the case of *want to*, when this form is orthographically transcribed into the colloquial variant, *wanna*, the spelling indicates that the form has been further reduced by omitting /t/ altogether while still retaining the voicing and nasality of /n/. The processes of reduction can, therefore, be represented as /ˈwɑnttu/ → [ˈwɑntu] →[ˈwɑntə] →[ˈwɑnə]. The decision to double the consonant in the spelling is likely due to the common spelling rule in English mentioned above in the *waddya* example.

The variant *ima* is a unique form, in which case three distinct processes of reduction are used to transform a four word phrase into one word. *Ima* derives from the phrase *I am going to*, and in terms of spelling, the processes of reduction can be shown as *I am going to* → *I'm going to* → *I'm gonna* → *ima*. In phonological terms, the initial contracted form *I'm* is the "result of PHONOLOGICALLY reducing a linguistic FORM so that it comes to be attached to an adjacent linguistic form" (Crystal, 1997, p. 89, capitals in original). The more frequent variant *gonna*, on the other hand, is the result of several effort-reducing processes. Initially, word-final /ŋ/ changes to /n/ influenced by the place of articulation from the conditioning sound /t/. Word-initial /t/ in *to*

is then simplified (or omitted), leaving the already assimilated form to retain the phonetic features of [n]. As Collins & Mees (2003) comment, when the form *going to* is used to form a tense, "it is pronounced /gənə/ in all but very careful speech" (e.g. *he's going to call* → /hiz gənə ˈkɔl/ (Collins & Mees, 2003, p. 212)). The form *gonna* is then further reduced and attached to the already contracted form *I'm*, represented by the orthographic transcription *ima*; phonologically speaking, the processes of reduction can be represented as /ˈaɪ æm ˈgoʊ ɪŋ tu/ → [ˈaɪm ˈgoʊ ɪŋ tu] → [ˈaɪm ˈgənə] → [ˈaɪmə].

Finally, in the case of the *–ing* suffix, word-final /ŋ/ is realized as [n]. Word-final <g> is omitted in the orthographic transcription of a variety of words (e.g. *havin, talkin, workin*, etc.). All of these variants are common to language use in EMC, and as Biber et al. (1999) note, variants of this type "[show] up in our data simply at the level of spelling" (p. 1122).

Additional Comments Regarding Spellings of Common Colloquial Variants

In the analysis of the nonstandard spellings of all the above examples, it will be necessary to draw upon the combination of processes. Certainly differences in dialect may influence differences in spellings, and not all EMC users necessarily conform to the more conventionalized forms of the informal colloquial variants. For example, some users represent /gənə/ as *gunna* instead of *gonna*.

Nevertheless, reasons to retain some (though certainly not all) of the spellings of the words in the assimilated and elided forms may be explainable in terms of phonology and morphology. First, as the phonemic and phonetic transcriptions show, the lexical words are those that are more stressed (a phonological characteristic common to lexical units (Biber et al., 1999)). In phonological terms, not only are stressed syllables less likely to reduce, but

word/syllable-final consonants are more prone to reduction than word/syllable-initial consonants (Shockey, 2003, p. 18).

With regards to spellings, part or all of the lexical words in several – though certainly not all – of the examples above have retained enough information for the reader to determine the underlying word(s), either from the spellings associated with particular phonemic realizations, or in the spellings associated with the form of the root morpheme of the word (*got(cha), wat(cha), wan(na), havin(g)*, *should(a), kind(a), im(a)*, and so on). At the level of morphology, then, it may occasionally be necessary to retain some representation of the root morpheme, since it is this feature that carries the most meaning in a word. As a result, in several of the examples, the word-final consonants do not always appear as they would in standard spellings, and, at times, letters are added or deleted, so as to represent a pronunciation of the colloquial variants. It is also worth noting that word-final <a> in all of the colloquial variants appears to be a representation of how the unstressed words like *you, to*, *of*, and *have* in particular sequences have been reduced to schwa [ə].

Finally, one possible reason why these specific colloquial variants are observed to occur more frequently in EMC (Anis, 2007; Palfreyman & Al Khalil, 2007; Tagliamonte & Denis, 2008) is perhaps due to their frequency in English: more frequent items are more likely to reduce (Shockey, 2003). Thus, the examples above are forms that correspond to particular colloquial and vernacular forms of spoken language, and are often reserved for informal writing and informal spoken conversations (Biber et al., 1999; Collins & Mees, 2003).

Additional Colloquial Variants

The shortening of words in writing (e.g. the word *night* is sometimes shortened to *nite*) is another well-documented feature in EMC and CMC research (Anis, 2007; Crystal, 2001 Herring,

2001). Anis (2007) calls this a "transcription of standard pronunciations" (p. 97), because it reflects the relationship between common spelling rules and the spoken pronunciation of the ideal form. The following variants will be considered here: *thru* and *nite*, which correspond to the standard written forms *through*, and *night* respectively. In the case of *thru*, it appears as though users are representing the sequence *ough* with the letter <u> to reflect its standard pronunciation in isolated form. For example, *through* is transcribed phonemically as /θru/, while the isolated name of the letter <u> is pronounced /ju/. Since *through* does not contain palatal /j/ in its standard realization, only /u/ is represented by <u>, and as a result, the informal, shortened variant is written as *thru*. So, in the formation of an alternative word, the spelling of *through* has been reduced to *thru* to more closely resemble a phonemic transcription of the word.

With respect to *nite*, it appears as though EMC users are attempting to reduce the form of the word, while still representing the ideal pronunciation in accordance with alternative common spelling rules. The ideal realization of *night* is transcribed phonemically as /naɪt/. Thus, the sequence *igh* in *night* (and other words) is ideally realized as the diphthong /aɪ/. In other English words, this diphthong also appears when word-final letter <e> forms "discontinuous composite vowel symbols" with, for example, the letter <i>, as in *site*; i.e. when the letters <i> and <e> are separated by one or more other letters, and when <e> is word-final, the spelling represents the diphthong /aɪ/ (Huddleston & Pullum, 2002, p. 17). So, if an EMC users wishes to represent /aɪ/ via alternative spellings, simplifying the *igh* sequence in *night* can be achieved by forming a discontinuous composite vowel with alternative spellings, in which case *nite* may be pronounced as /naɪt/. Furthermore, it is necessary to include word-final <e> so as to minimize confusion with a potentially different pronunciation. There is, for example, a difference between the pronunciations of *kit* and *kite* (realized respectively as /kɪt/ and /kaɪt/). Without the word-final

letter <e>, the diphthong <i> might be interpreted as /ɪ/, which could cause misinterpretation among participants.

Other Word Formation Processes

Abbreviations

As Delahunty and Garvey (2010) note, languages also have the potential to create new words by abbreviating. Abbreviation is one way to create other words by shortening already established words, and they are usually the informal variant of the original (Delahunty & Garvey, 2010, p. 136). The use of *comp* (whether in speech or in writing) instead of *computer* (or *composition, computation,* etc.), for example, occurs when the other syllables have been omitted. Often, the remaining syllable "provides enough information to allow us to identify the word it's an abbreviation of, though occasionally this is not the case" (Delahunty & Garvey, 2010, pp. 136-137). Abbreviations are highly frequent in EMC and are often one of the most remarked features of informal variants (Crystal, 2001).

Acronyms

Acronyms, on the other hand, are created by using the first letter of each word in a series of words or a set phrase (Delahunty & Garvey, 2010), such as *lol* (for *laugh out loud*), *brb* (for *be right back*) and *lmfao* (for *laughing my fucking ass off*); a maneuver all too familiar to EMC users, according to EMC scholarship (Anis, 2007; Baron, 2008; Barnes, 2003; Crystal, 2001; Crystal, 2008; Danet & Herring, 2007; Herring, 2001; Tagliamonte & Denis, 2008).

Blends

Blending occurs when select parts of two or more words fuse together to create a new word "whose form and meaning are taken from the source words" (Delahunty & Garvey, 2010,

p. 137). *Email* is a common example, which derives from *electronic* and *mail*, and which means a system for sending messages (i.e. *mail*) via electronic (the acronym of which is commonly referred to as *e*) telecommunication links between computer networks.

Rebus

Rebus writing is another type of writing that is commonly associated with EMC. In EMC, common forms of rebus writing include *c u l8r* (for *see you later*). Traditionally, rebus messages were defined as a form of writing that consisted entirely of pictures (or pictographs) to represent the sounds of words, instead of the objects to which they refer (Crystal, 2008, p. 39). Pronunciation classes and even party games (e.g. Dingbats) have made extensive use of rebus writing to demonstrate the relationship between sounds, pictures, and graphs (Crystal, 2008). For example, when a picture of two eyes is preceded by the letters *fr*, the sequence of the letters + picture can be read as *fries*. The following is another old example, which, apart from the last word, combines letter and numeral logographs (explained below) (taken from Crystal, 2008):

> YY U R YY U B I C U R YY 4 ME
> "Too wise you are, too wise you be, I see you are too wise for me." (p. 40)

Writing Systems and Terminology

Clearly EMC users are equipped with a repertoire of linguistic maneuvers to alter the writing system in their attempts to represent features of spoken language. Therefore, to simplify the selection of the sociolinguistic variables under investigation for the current study, I will present some terminology that, based on previous research, which will subsume all of the informal variants found in the current study into three overarching sociolinguistic variables. This section will begin with a discussion of the English writing system.

The term writing system is used here to describe the overall organization and structure of "graphically represented units of language" (Fouser et al., 2000, p. 53). A graph, therefore, refers to the smallest distinct visible segment in a writing system (Fouser et al., 2000, p. 53), such as the graphs *s, S, r, e,* and so on, as well as punctuation marks (Crystal, 1997, p. 176). Graphs can also be analyzed into graphemes (analogous to phonemes), in the sense that they are abstract contrastive graphic units in a script, which may be used to represent individual contrastive sounds (Crystal, 1997, p. 176; Crystal, 2005, p. 106). A script, therefore, refers to a set or collection of graphemes used together to form a stretch of writing or print (Crystal, 2005; Fouser et al., 2000, p. 53). The Roman alphabet is, thus, a collection of graphemes (commonly known as letters) that are used to write scripts in many writing systems across the globe, such as French, German, English, Turkish, etc.

Another important set of terminology refers to the various ways in which graphs are used in a writing system. Sampson (1985) made a distinction between systems of writing that are semasiographic and glottographic. The term semasiographic derives from the Greek word semasia, which refers to 'meaning' in the sense that the meaning of something can be graphically conveyed with the use of pictures, signs, or icons, rather than words (Boone, 1994, p. 14). According to Fouser et al. (2000), semasiographic systems of writing "represent concepts independently of the structures of language through icons;…[i.e.] a system of writing independent of spoken language" (pp. 53-54). On this interpretation, semasiographic systems convey meaning via the relationship between the icon and its intended referent (e.g. a road sign that depicts an image of a crosswalk represents that a particular location is where pedestrians are supposed to cross the street) (Boone, 1994, p. 16). Glottographic writing systems, on the other hand, refer to, at the very least, "sets of graphic signs corresponding to morphemic, syllabic,

and/or phonemic units which can be used to encode utterances through regular rules of sign combination" (Sampson, 1985, p. 29). That is, glottographic systems make use of visible segments (or graphs) to represent elements of a spoken language.

Sampson (1985) further divided glottographic systems into: (a) logographic systems, in which morphemes, lexical elements, and even parts of words are graphically represented; and (b) phonographic systems, which use graphs to represent the phonetic features in a language, such as syllables, sound segments, and features of articulation (e.g. the use of either *going to* or *gonna,* which also represent differences in pronunciation). According to Crystal (2005), this is a useful approach to interpreting writing systems, in that we can "classify them into cases that show a clear relationship between the symbols and sounds of the language (*phonological* systems) and those that do not (*non-phonological* systems)" (p. 108) While some non-phonological systems are primarily found in the early history of writing (e.g. cuneiform) (Crystal, 2005), the distinction between logographic and phonographic systems is the major division separating current writing systems in the world (Fouser et al., 2000).

Although it has been argued that a samasiographic system is technically impossible in the sense that speech must be reflected by all forms of writing (Unger & DeFrancis 1995), there is evidence in EMC and CMC research which shows that we can have a system of writing that is independent of the speech alone (Anis, 2007; Crystal, 2001; Crystal, 2008; Fouser et al., 2000; Herring, 2001). Though somewhat rare, semasiographic systems exist in airports (e.g. the arrow → sign to indicate direction to some location) and on highways (e.g. the picture of an airplane to indicate that by taking the next exit, the road will lead to an airport). In EMC, the pervasive smiley (e.g. :-)) is a prime example of how speakers of English can manipulate combinations of graphs to represent non-auditory, as well as nonverbal cues (Fouser et al., 2000, p. 54). EMC

users have implemented a wide array of alternative spellings and orthographic conventions to represent speech and other non-verbal and non-auditory cues in their electronic encounters.

The three distinctions to the writing system laid out by Sampson (1985) – phonographic, logographic, and semasiographic – form the three dependent sociolinguistic variables for this study. Exactly which specific features fall into each variable category will be discussed in the methods chapter that follows. The frequency of use of the three sociolinguistic variables will help answer the two research questions in the current investigation of language use in EMC:

1. How does age correlate with the linguistic behaviors of EMC users when they use informal variants to represent features of speech and orality in the transmission of ETMs?
2. How do EMC users' perceptions of formality influence their decision to represent features of speech and orality with the use of informal variants in electronic text-based interactions?

The following hypotheses have been developed in response to the research questions set out above:

1. EMC users use a variety of informal variants to represent speech and orality, but the extent to which these variants are used will vary according to the age of the participant. Specifically, I expect to see a higher frequency of use of informal variants among younger age groups, and that this frequency will decline as the age of the participants' increases.
2. The use of informal variants to represent speech and orality in EMC will vary according to the individual participant's perception of the level of formality.

Specifically, I expect the ETMs which the participants identify as formal to contain less frequent use of informal variants, and the reverse for informal contexts.

Chapter 5: Methodology

Data Collection

Participants

Three types of data were gathered from 33 participants - age specific demographic information, text message data and data from email correspondence. The participants were categorized into three distinct age groups. Participants were assigned to three discrete age ranges for three reasons: (1) there is, according to Ling (2005), an overall tendency for individuals below the age of 24 to show a higher frequency of use of informal variants whenever they interacted with others electronically; (2) it is a common practice in EMC research to place an age cut-off point no higher than 25, since the assumption is that "regular users [of EMC] are typically members of younger generations" (Anis, 2007); and (3) it is necessary to keep the age range intervals consistent among the group of participants in the study. Participants were, thus, divided along the same seven-year age range intervals (i.e. 18-24, 25-31, 32-38). Table 1 summarizes mean ages and standard deviations.

Table 1
Means and Standard Deviations of EMC users indexed by age (*N* = 33)

Age (years)	*N* size	Mean	Standard Deviation
18-24	12	22.58	1.16
25-31	11	27.63	1.69
32-38	10	33.7	1.57

Although convenience sampling procedures were used in the recruitment process of the overall sample, the final selection of the participants was based on a procedure more comparable to that of a stratified random sampling. That is, although participants were recruited on a voluntary basis, the final selection of the participants relied primarily on age specific criteria gathered from the completion of a preliminary demographic survey. This enabled the sample to

consist of participants who were categorized into the three distinct age groups. Informed consent was gathered from all participants and all their names were anonymized. Furthermore, it is worth noting that the participants in this study were invited to attend a small dinner party, as compensation for their participation.

To ensure the homogeneity of the EMC users' backgrounds in this study, all participants were selected on the basis of two criteria: (1) all participants assured the researcher that they were native speakers of American English; and (2) all participants had received some form of higher education in the US. The former criterion enabled the researcher to draw conclusions about variations in language use from this group of English EMC users in particular, whereas the latter was necessary for the sake of maintaining a relatively uniform level of education among the participants.

Survey

A preliminary survey was used to gather demographic information from the participants, (see Appendix A). Participants were asked to complete a survey prior to their submission of the data. A total of five items were included in the survey. Participants were asked to identify their age, level of education, and affiliation with Colorado State University, as well as confirm that they had an email account and a mobile phone that could transmit text messages. The survey's primary purpose was to yield a set of age specific demographic data that could be measured against the linguistic features found in the corpus (outlined below). Once consent was obtained, participants were emailed two documents: (1) an electronic version of the survey to complete and send back to the researcher; and (2) a set of instructions informing them about how to submit electronic text-based data for the compilation of an EMC corpus (see Appendix B).

Procedure: Compiling the Corpus

In this study, the participants contributed 2,542 message transmissions, resulting in the compilation of a 47,739 word corpus. The corpus was generated by asking participants to provide authentic ETM data. Because previous studies (Baron & Ling 2007; Ling, 2005) that have analyzed the distribution of variable linguistic features found in EMC, data for the present study was gathered along similar lines of methodology. Participants were asked to submit messages they transmitted over a two week period prior to the data gathering period. Rather than gather the data over an observed time interval, as has been done in previous studies (Carr, Schrock & Dauterman 2012), requesting data from messages the participants had already created would hopefully mitigate the observer's paradox dilemma described by Labov (1972), insofar as the researcher had no influence on the language in their messages – though participants did have the option to select the messages they wanted to include. Participants were also asked to submit only the messages they produced, so as to protect the identity of their interlocutors, who had not given consent to participate in the study. Participants were assured that both their personal information and in the contributions would remain confidential. To ensure confidentiality, each participant was assigned a random code (e.g. 1012) that linked to his or her data. All other names in the data were replaced with an underlined space (e.g. _____).

The data came from two sources of EMC – email and text messages. Both forms of EMC are types of private correspondence; however, according to some researchers (Baron, 2008; Crystal, 2001; Herring, 2001; Danet & Herring, 2007), the two media are believed to be separated along the continuum of synchronicity (p. 7 in this study). The common perception is that less synchronous forms of EMC (e.g. email) are likely to display more standard (and in effect, more formal) language use than more synchronous forms (e.g. text messages), due to time

constraints that limit users' ability to edit the messages they compose (Baron, 2008; Crystal, 2001; Herring, 2001). Although this was not the focus of the current study, the relationship between the two media and the linguistic features was also calculated, from which possible implications may be drawn.

Text Messages

A common protocol used in mobile phone studies (Baron 2008) is the collection of a corpus by distributing a template for participants to write diaries that record exactly all the text messages they transmit over an observed period of time. However, with the development of more sophisticated telecommunications systems, many of the participants were already equipped with technological resources and/or functions on their mobile phone devices to send archived messages to the researcher directly via email. Since this proved to be a more efficient method that additionally limited the number of transcription inaccuracies, this was the preferred technique (though the researcher still provided participants with a transcription template sheet shown in Appendix B, in the event that some technological constraint prevented participants from transmitting ETMs electronically).

Emails

Like text messaging, email has established itself as a primary means of interpersonal communication with the communicative function of constructing and maintaining social relationships (Barnes, 2003) as well as transferring information. Email is, therefore, another important medium in which to observe how EMC users manipulate their writing in their online encounters. Participants were asked to copy and paste only the emails they composed during the two-week time interval, and to forward them to the researcher.

Data Coding and Analysis

To determine the extent to which the participant's perceptions of formality influence language use in EMC environments, participants were asked to identify the level of formality with the use of two codes – formal (F) and informal (I). Allowing the participants to identify levels of formality not only eliminated any value judgments imposed by the researcher, but also enabled them to personally identify the nature of the context in which the message was transmitted, a component that is critical when interpreting stylistic variation in sociolinguistics. These codes were used as the second set of independent variables that enabled the researcher to determine how linguistic variation correlates with stylistic differentiation. For uniformity of analysis, participants were asked to identify whether the messages they submitted were: (1) a text message (TM); or (2) an email (E).

The Sociolinguistic Variables

Drawing upon previous research dealing with the frequency of use of phonographic, logographic and semasiographic features (Fouser et al., 2000), the present study follows a similar framework for the selection and classification of the sociolinguistic variables (the dependent variables). Investigations exploring the language of EMC suggest that users often tend to use colloquial variants (e.g. *wanna, gonna, watcha*, etc.), and/or shorten words (e.g. *thru, nite, diz,* etc.) to represent features of speech and pronunciation, as well as lengthen words (e.g. *sooon*, *yaaay*, etc.), represent laughter (e.g. *haha*, *hehe*, etc.), and use capitals (e.g. *VERY*, *GREAT*, etc.) and repeat punctuation marks (e.g. *why???*, *well…*, *no!!!*, etc.) to add emphasis (Anis, 2007; Barnes, 2003; Baron, 2008; Crystal, 2001; Crystal, 2008; Fouser et al., 2000; Herring, 2001; Palfreyman & Al Khalil, 2007; Squires, 2010; Tagliamonte & Denis, 2008). All instances of these linguistic varieties will be classified as phonographic features. Other maneuvers to

represent speech include the use of single letters (*b* for *be*), numerals (*4* for *for*), typographic symbols (@ for *at*), as well as abbreviations, acronyms, and blends (e.g. *omg, lol, brb*, *frolf*, and so on) (Barnes, 2003; Baron, 2008; Crystal, 2001; Crystal, 2005; Crystal, 2008; Fouser et al., 2000; Herring, 2001). All instances of these linguistic varieties were classified as logographic features. Finally, the use of emoticons and other nonverbal, non-auditory cues were classified as semasiographic devices. Examples from coding of the data are illustrated in Table 2. Results from the coding of the data are shown Appendix C. (see Appendix D for a complete list of informal variants found in the corpus of ETMs).

Table 2
Examples and Classification of Features

Type	Definition	Examples
Phonographic	graphs used to represent the phonetic features in a language, such as syllables, sound segments, and features of articulation	SOOON haha ya gonna…nite
Logographic	graphs used to represent morphemes, lexical elements, and parts of words	c u fml lol omw @ b/c btw
Semasiographic	graphs used to represent concepts not directly tied to spoken language	:) :/ <3 ;) :p B-) ^ ^ :-*

All the participants were categorized into three discrete age groups. The different media were then separated into the two categories (email and text messages), and then subcategorized according to the level of formality identified by each participant. For reliability, a second coder was given 10% of the corpus to examine, resulting in an inter-coder agreement of 96.8%. Raw counts of instances when an EMC user produced informal variants were counted and separated

into the three dependent variable classifications outlined above. Frequency of use of all instances of the three dependent variables were tallied and their raw counts were used to test the effect of the first independent variable: age. Raw frequency counts were analyzed using Pearson's *r* product-moment correlation coefficient to assess the degree to which the quantitative variables are linearly related.

In order to test for correlations between the dependent variables and the second independent variable (formality), however, all raw counts were normalized according to common normalization processes. "'Normalization' is a way to adjust raw frequency counts from texts of different lengths so that they can be compared accurately" (Biber, Conrad, & Reppen, 1998, p. 263). The raw numbers of linguistic features under investigation were thus divided by the total number of words in the text and then multiplied by an appropriate norming standard (in this case, by 1000), specifically because the sizes of the discrete corpora in relation to formality were unequal in size: in every single age group, formal text message corpora amounted to less than 200 words, whereas informal text message corpora amounted to more than 3,000 word tokens. Normalization was, therefore, seen as a necessary procedure to avoid misrepresentation of correlate values by way of maintaining more equally comparable counts in proportion to the size of each corpus. Normalized data counts within each media were then entered into SPSS Point biserial correlation test to determine the effect of formality.

Chapter 6: Results

Distribution of Features across Age Groups

In order to determine how age correlates with the linguistic behaviors of EMC users when they use informal variants to represent features of speech and orality in the transmission of ETMs, raw counts were used to determine the distribution of features of language use across age groups. Overall frequency of use of features used to represent speech and orality in EMC in relation to age is shown in Table 3.

Table 3
Raw Frequency Counts of Each Linguistic Variable by Age Groups

Age group (years)	Phonographic	Logographic	Semasiographic
(18-24)	836	53	99
(25-31)	546	50	34
(32-38)	344	195	129
Total	1726	298	262

The first hypothesis proposed the following:

1. EMC users use a variety of informal variants to represent speech and orality, but the extent to which these variants are used will vary according to the age of the participant. Specifically, I expect to see a higher frequency of use of informal variants among younger age groups, and that this frequency will decline as the age of the participants' increases.

The data in Table 5 reveal that the first hypothesis was only partially supported. Raw counts indicate that as age increases, the use of phonographic features to represent speech and orality in EMC occur less frequently. (Note: the phonographic features occurred more frequently in the overall corpus, regardless of how the data were divided according to the two independent variables). As for logographic and semasiographic features, the oldest age group produced these

features more frequently than any other group. Only among members of the two younger age groups does the hypothesized pattern between age and these features emerge from the raw scores, yet the difference in frequency of use of these two variables and the two younger age groups is not nearly as dramatic. Figure 1 shows a graphical representation of the data in Table 5.

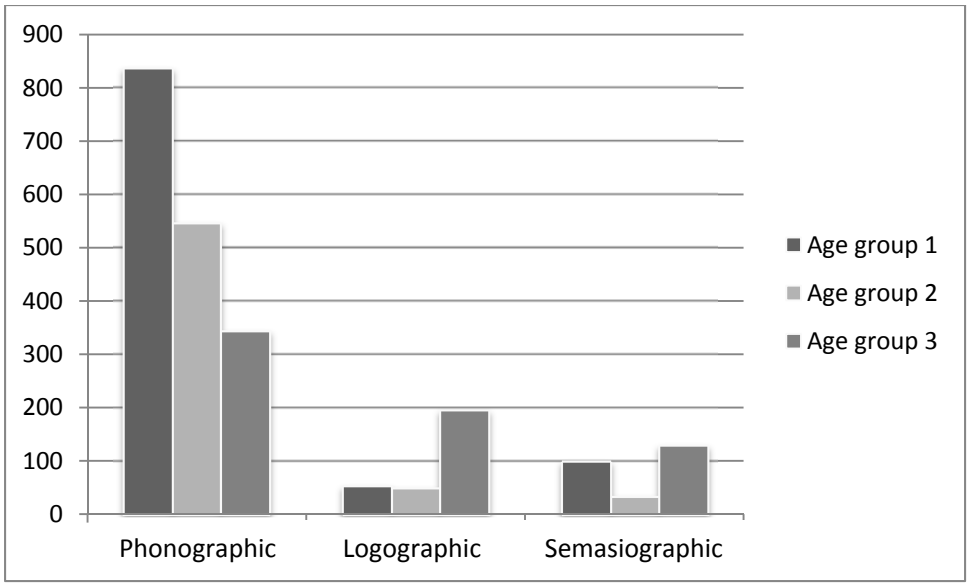

Figure 1. Raw frequency counts of each linguistic variable by age groups

Correlations were calculated to determine the distribution of features across age groups. Analysis of Pearson's *r* product-moment correlation also only partially supported the hypothesized relationship between age and the three dependent variables. In terms of age differentiation, the results show a weak inverse relationship between the frequency of use of phonographic features to represent speech and an increase in age ($r = -.23$). A weak positive relationship was found for frequency of use of semasiographic features and an increase in age ($r = .12$). Significant positive correlations were only found between logographic features and age ($r = .39$, $p < 0.05$). The results indicate that the older the participants were, the more likely they were to use logographic features in EMC. An unexpected significant positive correlation also

emerged regarding the co-occurrence of semasiographic and of logographic features. Analysis of the data revealed that participants were more likely to produce one or several instances of these features whenever one or several instances of the other feature appeared in the corpus of ETMs ($r = .39$, $p < 0.05$).

Distribution of Features across Formal and Informal Contexts

In order to determine the extent to which EMC users' perceptions of two levels of formality influence their decision to use informal variants to represent speech and orality with the use of informal variants in electronic text-based interactions, normalized data counts were used to determine the distribution of features across formal and informal contexts of language use. Overall frequency of normalized counts and the distribution of features across formal and informal contexts of language use are shown in Table 4.

Table 4
Normalized Frequency Counts of Each Linguistic Variable by Formality (per 1000 words)

	Phonographic	Logographic	Semasiographic
Formal	13.16	3.09	4.32
Informal	155.17	26.10	21.80
Total	168.33	29.19	26.12

The second hypothesis proposed the following:

2. The use of informal variants to represent speech and orality in EMC will vary according to the individual participant's perception of the level of formality. Specifically, I expect the ETMs which the participants identify as formal to contain less frequent use of informal variants, and the reverse for informal contexts.

The data in Table 6 show that overall normalized counts support the second hypothesis. The data reveal a marked difference between frequency of use of phonographic, logographic and

semasiographic features and the level of formality. The data also indicate that the contexts in which the participant perceived were more formal, the less likely they were to use the features under investigation. Figure 2 illustrates the normalized counts and the distribution of features across formal and informal situations of language use.

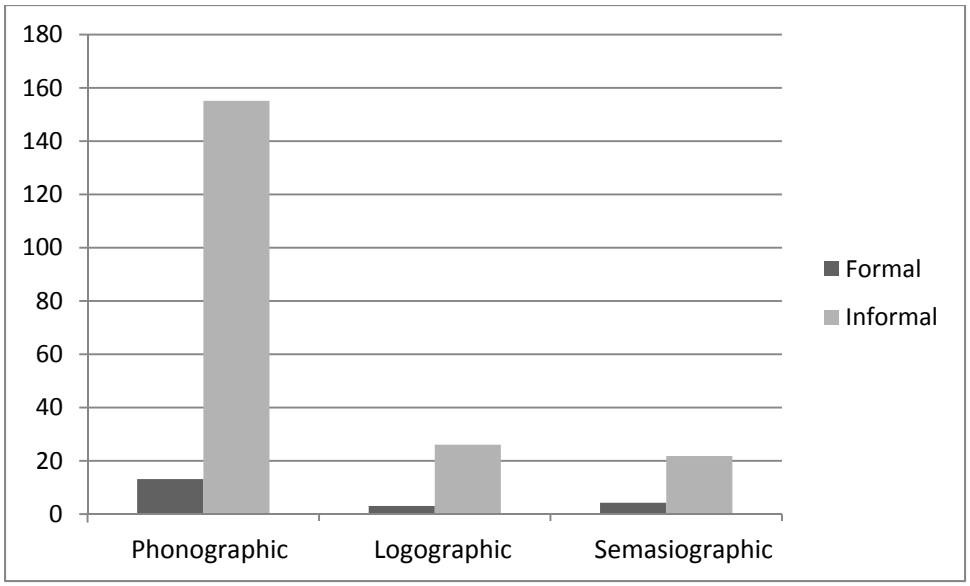

Figure 2. Normalized data counts of each linguistic variable by formality.

Correlations were calculated to determine the distribution of features across formal and informal contexts of language use. The results indicate that the hypothesized relationship, particularly between formality and phonographic features, was confirmed. The use of phonographic features to represent speech were the most frequently used variable in the current study of EMC, and there was a significant inverse relationship between formal contexts and the use of these features (r = -.86, $p < 0.05$). Though not significant at the $p < 0.05$ level, there were strong inverse correlation patterns between formal contexts and the frequency of use of logographic and semasiographic features (r = -.67 and r = -.81, respectively).

Chapter 7: Discussion

As stated in the introduction, the present study is a sociolinguistic interpretation of language variation in EMC in relation to two commonly investigated extralinguistic factors: age and formality. Analysis of the data revealed several relationships between these factors and the sociolinguistic variables identified by the researcher. The discussion will begin with an examination of how age correlates with language variability in EMC, followed by an interpretation of the effect of formality.

Prior to presenting an analysis of the data, a brief mention of how the results were interpreted is necessary. Once the data were entered into the two correlation tests, a level of significance was established at $p < 0.05$ for all correlation analyses, and the results were interpreted according to the guidelines laid out by Salkind (2010). Based on Salkind (2010), a common method used to assess the strength of the relationship between the variables in the population is to determine where the index of the effect size lies along a linear range of values. Naturally, the index value ranges from -1.00 to +1.00, and the higher the value of the correlation (i.e. the higher the *r* value) is within this range, whether positive or negative, the stronger the relationship. To aid in interpreting the strength of the relationship between variables, I will refer to the strength of the *r* values as they fall within the indices presented in Table 5 (taken directly from Salkind, 2010, p. 88):

Table 5
Value Ranges Used for Interpreting the Correlation Coefficient

Size of the Correlation Coefficient	*General Interpretation*
0.8 to 1.0	Very strong relationship
0.6 to 0.8	Strong relationship
0.4 to 0.6	Moderate relationship
0.2 to 0.4	Weak relationship
0.0 to 0.2	Weak to no relationship

The Effect of Age

Regarding the first research question, which queried how age correlates with the linguistic behaviors of EMC users when they use informal variants to represent features of speech and orality in the transmission of ETMs, the results show that age only correlated significantly ($r = 0.39$) with the use of logographic features at the $p < 0.05$ level. Raw scores in Table 3 show that the oldest age group used these features more frequently than all other age groups in the current study. As for phonographic and semasiographic features, no significant correlations were found. Although the raw data show a decline in use of phonographic features as the age of the participants' increases, analysis of the data reveals that age is not a significant correlate in the overall distribution of phonographic features. With regard to semasiographic features, the raw data show a pattern which revealed that members of the oldest age group produced these features more frequently, followed by members of the youngest age group, and then by members of the middle age group. Still, no significant correlations were found between semasiographic features and age. There was, however, a relatively moderate significant correlation found between the use of semasiographic and logographic ($r = 0.39$) features at the $p < 0.05$ level. Analysis of the data reveals that these features are likely to co-occur in EMC language when distributed across age groups.

One aim in this study was to determine the extent to which age differences relate to the linguistic practices of EMC users. Often, younger users, particularly teenagers and young adults, are associated with the use of informal variants considered unacceptable or inappropriate for formal written texts, including the use of certain features which the current study subsumed under logographic features (e.g. acronyms, abbreviations, etc.) (Crystal, 2001; Danet & Herring, 2007; Squires, 2010; Tagliamonte & Denis, 2008). With respect to the significant correlations

found between age and logographic features, the results show an opposite trend. In the current study, older age groups used logographic features more frequently in EMC environments than any other age group, and this finding is further supported by the raw scores – individuals between the ages of 32-38 produced nearly half (129 of the total 262 instances) of all the logographic features in the corpus.

One possible explanation would be to interpret the more frequent distribution of the logographic features as a way for the older participants to mark group membership in the current corpus of ETMs. That is, the significant correlations found between logographic features and members of the oldest age group may be interpreted as a preference by older users to retain certain features that they may have acquired in their earlier experiences with EMC. More than a decade ago, Crystal (2001) claimed that the various types of abbreviations that appear in EMC were "one of the most remarked features" of the time (p. 84). Analysis of EMC data gathered one years later in Ling's (2005) study led him to claim that "in support of the popular stereotype, however, teens and young adults SMS seem to be the biggest users of abbreviations" (p. 344). The statistical significance found between abbreviations and EMC users' ages at the $p < 0.001$ indicated to Ling (2005) that the use of these features by teens and young adults in 2002 contributed to a sense of group membership in relation to the age differentiators that set them apart from other age groups.

Similar findings were documented elsewhere (Baron 2004). In 2003, Baron (2004) collected a corpus of EMC messages from 22 American college students to analyze language use in IM and text messages. In her analysis of several linguistic features, Baron (2004) found that acronyms, particularly the use of *lol*, were commonly used "as a phatic filler; roughly comparable to *OK, really*, or *yeah* in spoken discourse" (p. 416).

In the current study, the significant correlations found between logographic features and members of the oldest age group may indicate that older age groups retained the widespread use of logographic features from the time when these were the most documented features of EMC language. A decade ago, the EMC users in the current investigation would have matched the young adult age intervals examined by Ling (2005) and Baron (2004). So, the significant correlations found between the use of logographic features and members of the 32-38 age group may indicate that older users may have retained the use of logographic features until present day, so much so, that they now outperform all other age groups in the distribution of these features.

With regards to the overall distribution of semasiographic and phonographic features in relation to age, no significant correlations were found. Whereas an increase in age correlated positively with the use of semasiographic features (r = .12), an inverse correlation was found between an increase in age and the use of phonographic features (r = -.23). At first glance, the positive and negative correlations indicate that the younger an individual was, the more likely they were to use phonographic features, and the reverse for semasiographic features in this corpus of ETMs. Analysis of raw scores might actually confirm this claim regarding the use of phonographic features in particular: members of the youngest age group produced 836 of the total 1726 instances (48%) of phonographic features identified throughout the corpus, followed by members of the middle age group who produced 546 of total instances (32%), and then by members of the oldest age group who produced 344 of the total instances (20%). Yet, the weak inverse correlation accompanied by a lack of statistical significance appears to indicate that phonographic features are not necessarily associated with age. Instead, analysis of the data indicates that, at least among the group members in the current study, EMC users within the 18-38 age range may use phonographic features to represent speech and orality to a lesser or greater

extent, though the extent to which they use phonographic features is not dependent on the age of the individual. The implication is that representing speech and orality in EMC via the use of phonographic features is not merely a teenager practice.

A similar claim might be made in terms of the frequency of use of semasiographic features. The results indicate that the use of semasiographic features increases as the age of the participants increases, though without any statistical significance. Simply in terms of raw data, members of the oldest age group used semasiographic features more frequently, especially in the case of the prototypical smiley face. Not only did members of the oldest age group produce this feature 99 of the total 193 times (over 50%) in the entire corpus, but it was only among this group of participants that the smiley face, alongside other emoticons, frequently occurred in formal contexts. Again, however, weak, positive correlations, as well as the lack of statistical significance, suggest that these features are not necessarily associated with the age of the participants.

The fact that neither phonographic nor semasiographic features correlated significantly with the age of the individuals brings up an interesting point. Not only do these findings indicate that age is a relatively weak factor among the current group of EMC users, but it also poses the question regarding whether or not the widespread use of these particular features is becoming acceptable for EMC language among individuals in the 18-38 age range. If Labov (2006) is correct in his interpretation of language change, in which he argues that the overt acceptability of vernacular terms may be described "as a *change from below*, because it expresses a gradual shift in the behavior of successive generations, well below the conscious awareness of any speaker" (p. 206) then it is possible that the use of, specifically, informal colloquial variants and emoticons may have spread to the extent that all age groups now use these informal variants

(though some participants use these features more than others), resulting in the non-significant relationship between age and the use of phonographic and semasiographic features. As the following examples illustrate, phonographic and semasiographic features were used by members of each age group (example 1 comes from an EMC user in the 18-24 age range; 2 from a user in the 25-31 age range; and 3 from a user in the 32-38 age range) in the current study, though the distribution of these features varies according to the individual – phonographic features are underlined, whereas semasiographic features are marked in **bold**:

1. *You're gonna be such a beautiful bride ☺send me pix if you need opinions!*
2. *Really don't care **:)** I'll take em ya comps mama unless you need em like right now*
3. *Hey chica! Yep – I think we should start planning soon. should be better after the 22 and we can make some decisions then. ☺*

Lack of sufficient sociolinguistic research in the area of EMC makes it difficult to substantiate this claim; however, the fact that Ling (2005) found younger EMC users to be the group most prone to use what the current study categorized as phonographic and semasiographic features in a 2002 corpus of EMC language may indicate that the spread of these specific features has changed the linguistic behaviors of successive generations to adopt these forms in their electronic interactions.

Finally, there was an unexpected finding regarding the use of semasiographic and logographic features. Analysis of the data revealed that these features are more likely to co-occur in the current corpus of ETMs with a statistical significance of $r = 0.39$ ($p < 0.05$). Consider, for example, the following message transmitted by a member of the oldest age group:

4. *Ok good ;) well then I'll c u after school*

In this message, three of the ten tokens are either semasiographic (e.g. *c* for *see*, *u* for *you*) or logographic features (considering that ;) is regarded as one token). The tendency to use both semasiographic and logographic features in the same message occurred quite frequently

throughout the corpus, especially among members of the oldest age group. This tendency may be considered one reason for why there was an overall increase in the frequency of use of these features by members of the oldest age groups in the raw scores. It is likely that the older EMC users in the current study prefer the use of both features when they communicate electronically.

The Effect of Formality

A second goal of this study was to determine the extent to which differences in formal and informal contexts influence the linguistic practices of EMC users. In just about every sociolinguistic investigation that has observed the stylistic dimension of language variation, the degree of formality appears to be the most significant factor that shapes language use (Coupland, 2007). As outlined in chapter 2, the level of style is considered the most widely available approach to observing variation for (monolingual) speakers, and in the sociolinguistic context, stylistic variation "refers to speech variations that reflect one's assessment of the social context and of what is or is not 'appropriate'" (Edwards, 2009, p. 28). Unsurprising, then, is the fact that the degree of formality is the most common influence and product of how one shifts their linguistic behavior according to the context (Edwards, 2009). The present study confirms these sociolinguistic trends.

In response to the second research question asking how EMC users' perceptions of formality influence their decision to represent features of speech and orality with the use of informal variants in ETMs the results show that there was a strong, significant inverse correlation (r = -0.86) between the use of phonographic features and formal contexts at the p <0.05 level. Thus, phonographic features were frequently reserved for informal contexts by all participants in the current study. Although no statistical significance was found between formality and the use of logographic and semasiographic features, analysis of the data revealed

that both features exhibit strong inverse correlations (r = -.67 and r = -.81, respectively) with contexts that the participants identified as formal. On the whole then, the data indicate that the participants were less likely to use any of the nonstandard linguistic features in contexts which they perceived to be more formal.

When asked to assess the formality in the messages, the participants almost exclusively identified messages that contained more frequent instances of the dependent variables as messages produced for informal contexts. Most notable was the frequency of use of phonographic features, which appeared to be the sociolinguistic variable that showed the greatest variation with respect to formality. The significance of this variable as a correlate of stylistic variation was found at the $p < 0.05$ level. The distribution of phonographic variables across formal and informal contexts suggests that participants prefer to reserve the use of phonographic features for contexts they perceive to be informal. As Danet and Herring (2007) have argued, the frequent use of features of this sort show "a tendency toward speechlike informality" (p.12). As one participant mentioned, as sort of a disclaimer when submitting messages to the researcher, "just so you know i try to fashion my informal texts into something resembling a spoken slang…disregarding the conventions of standard english…i try to use verbal shortcuts in informal texts." Simply in terms of normalized data counts, the pattern becomes more obvious, as the numbers show that participants produced phonographic features at least ten times more often in informal contexts than formal ones.

With respect to logographic and semasiographic features, on the whole, the participants in this study tended to reserve these features for more informal contexts, though no significant relationship was found at the $p < 0.05$ level. There were, however, several instances in which the use of logographic and semasiographic features occurred in contexts that the participants

identified as formal. In email correspondence, for example, the middle age group produced more instances of logographic features in formal contexts than in informal ones. In terms of normalized counts, members of the oldest age group produced a greater number of semasiographic and logographic features in formal text messages than in informal ones. Possible explanations for the lack of statistical significance found between formality and the use of logographic and semasiographic features might include: (1) the use of logographic (especially acronyms and abbreviations) and/or semasiographic features are, at least in part, linguistic variants becoming somewhat acceptable for formal contexts; because (2), their greater frequency of use by members of older age groups may be an indicator that these forms are gradually becoming conventionalized for both formal and informal contexts.

It is difficult to evaluate the generalizability of these claims, due to the limited number of participants, as well as the fact that the participants were only provided with two levels of formality when asked to codify the messages (explained in the limitations below). Nevertheless, it is safe to say that the contexts in which speakers produced more instances of each dependent variable were identified as informal; keeping in mind, however, that the participants were allowed only two levels of formality. The social meanings attached to these linguistic forms may, as a result, be linked to the particular context in which the participants interacted (Johnstone, 2010). This is because the norms associated with the linguistic features as they relate to the particular interactional context are evidence that several of the linguistic forms used to represent speech and orality by EMC users are typically reserved for informal communication, though not entirely, as in the case of older group members.

Limitations

Data Gathering

The current study entailed a number of limitations in the data gathering approach. First, the participants in the study were recruited on a voluntary basis, and for a few participants, their involvement in the study was contingent upon compensation for their services. Thus, the ETMs used in the compilation of the corpus were only gathered from participants who were willing to share their messages, thereby limiting the generalizability of the findings.

Second, in most cases it was only possible to gather messages that the participants had sent themselves. Although some participants were capable of submitting messages they sent and received, so long as their interlocutor(s) had also signed the consent form, this occurred very infrequently. For the majority of the participants, the researcher was not able to gather messages the participants had received for both ethical and methodological reasons. Ethically, it was not possible to request messages that an EMC user had received, since doing so would have included data from people who had not given consent. There is also the methodological problem that the demographic information of the senders would be unknown to the researcher. Without such information, it would have been impossible to analyze the data meaningfully.

Third, the messages gathered by the researcher were only those which the participants felt comfortable contributing to the research. Therefore, it is unlikely that the data are fully representative of the types of language use practiced and shared by the EMC users in the current study. On the one hand, the researcher attempted to mitigate the observer's paradox by asking participants to submit messages they had produced prior to giving consent. On the other hand, however, much like Rickford and Eckert's (2001) critique of the sociolinguistic interview, the types of language that the researcher wished to observe may not have been provided for current research, due to issues of privacy or lack of a secure sense of confidentiality. Concerns about

confidentiality were some of the commonest reasons why some individuals preferred not to participate in the study. As Coupland (2007) states, the authenticity of a speaker's linguistic behavior is not always observable in sociolinguistic inquiries, since quite often there is no established intimate relationship between the participant and the analyst. The participants were therefore compelled to decide which messages they felt were appropriate to contribute to a corpus for research and which messages they did not want the researcher to analyze.

A further weakness of the data gathering techniques in the present study is that the messages were often taken out of context. Although the messages were transmitted for particular communicative purposes in a series of interactional contexts, the influence that another person had on the linguistic behaviors of the participants was impossible to control for. Ling (2005) also struggled with this methodological limitation by concluding that "it is difficult to estimate the degree to which this is an issue given the stricture against examining incoming messages" (p. 337). In general, however, the decontextualization prevented the possibility of pursuing a meaningful discourse analysis of the features.

Unequal Distribution

Another limitation to the study was the unequal distribution of messages contributed to the corpus by the participants. Not every participant contributed the same number of messages and words, nor did they contribute an equal amount of formal and informal and/or email and text message data. As mentioned in the discussion above, only seven participants identified a very small portion of text messages as formal. Additionally, there was an asymmetrical distribution of formal and informal email messages by the participants in the study. Some individuals contributed well over 2000 word tokens for research, whereas others provided less than 500

word tokens. As a result, some individuals exhibited greater linguistic variability in comparison to others, simply because they submitted a higher word count to the overall corpus.

Number of Contexts

Finally, it is also worth noting that the participants were only given two options from which to determine the level of formality in their messages. It was, therefore, not possible to set up a hierarchical analysis of style, in order to determine the extent to which several distinct levels of formality might influence variable distributions of language use in EMC. Based on the sociolinguistic observations presented in chapter 2, it is possible that certain features would have been used more or less frequently according to several different levels of formality. It is also likely that the distribution of features may have paralleled similar patterns documented in previous sociolinguistic research to the extent that the results might have shown a steady progression in some direction, gradually increasing or decreasing according to the particular level of formality under analysis.

Conclusion and Implications

Although ETMs of participants affiliated with an American university in the 18-38 year age range in the present investigation display considerable variation of language use across all age groups, as a whole, they evidence that the practice to represent speech and orality with the use of phonographic and semasiographic features is not necessarily a teenager phenomenon. Analysis of informal linguistic variants in the current corpus of EMC language suggests that only logographic features can be associated with age, a finding which can best be understood in the real-time hypothesis of language variation and change. Comparison of results found in this study with those found in previous sociolinguistic research examining language use in EMC (Anis, 2007; Baron, 2004; Ling, 2005; Palfreyman, & Al Khalil, 2007) indicates that members of the

oldest age group in the current study have likely retained certain linguistic features – abbreviations and acronyms – which were found to correlate significantly with members of corresponding age groups approximately ten years earlier. It would be interesting to see if future research can confirm whether these linguistic features remain to be a marker of group membership by older EMC users.

Consistent with previous sociolinguistic research, the current study found that each nonstandard linguistic feature under investigation correlated negatively with contexts that the participants identified as formal. As was predicted, the use of informal variants in ETMs is more likely to appear in EMC environments that are considered informal by the participants. The extent to which these features are distributed across a broader range of levels of formality is an area that remains to be studied. Nevertheless, the significant inverse correlations found between phonographic features and formal contexts indicate that the decision to represent speech and orality by altering traditional spellings and orthographies is a practice reserved for informal contexts.

There were also two unexpected findings regarding to the co-occurrence of semasiographic and logographic features across age groups. The significance of this finding indicates a common practice by EMC users in the current study, implying that participants are likely to transmit ETMs with logographic features accompanied by semasiographic features. Whether or not this practice can be understood as a way to provide some frame of discourse, or whether the frequent use of these features is simply a type of phatic filler in electronic text-based communication as indicated by Baron (2004), additional research is needed to determine the extent to which this phenomenon recurs among other EMC users.

To study language variation in EMC seriously, empirical research needs to incorporate a number of delimited studies, focusing on level of education, age, gender, as well as on other factors in the field of sociolinguistics, such as the communication objective, audience design, and so on. The limited number of previous investigations observing language use in EMC from a sociolinguistic point of view clearly restricts the ability to make generalizable interpretations about the linguistic behaviors of EMC users as a whole. Thus, the scope of research should be enlarged to observe which other extralinguistic factors tend to influence language use in EMC. Above all, much research remains to be conducted regarding spelling and orthographic variations in EMC, in order to confirm or disconfirm the trends posited here.

References

Anis, J. (2007). Neography: Unconventional spelling in French SMS text messages. In B. Danet & S. Herring (Eds.), *The multilingual internet: Language, culture and communication online* (87-115). Oxford, UK: Oxford University Press.

Anttila, A. (2002). Variation and phonological theory. In J.K Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change*. (206-243). Malden, MA: Blackwell Publishers Ltd.

Axtman, K. (2002) "r u online?: The evolving lexicon of wired teens. *Christian Science Monitor 12.* Retrieved from http://www.csmonitor.com/2002/1212/po1so1-ussc.html.

Baron, N. (2004). See you online: Gender issues in college student use of Instant Messenger. *Journal of Language and Social Psychology, 23*. (397-423).

Baron, N. & Ling, R. (2007). Text messaging and IM: Linguistic comparison of American college data. *Journal of Language and Social Psychology*, *26.* (291-298).

Baron, N. (2008). *Always on: Language in an online and mobile world*. Oxford, UK: Oxford University Press.

Barnes, S. (2003). *Computer-mediated communication: Human-to-human communication across the internet.* Boston, MA: Pearson Education, Inc.

Becker, K. (2009). /r/ and the construction of place identity on New York City's Lower East Side. *Journal of Sociolinguistics,13*, (634–658).

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use.* Cambridge, UK: Cambridge University Press.

Biber, D., Johannsen, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow, England: Longman.

Boone, E. H. (1994). *Writing without words: Alternative literacies in Mesoamerica and the Andes*. Durham, NC: Duke University Press.

Bucholtz, M. & Hall, K. (2010). Locating identity in language. In C. Llamas & D. Watt (Eds.), *Language and identities*. (18-28). Edinburgh, UK: Edinburgh University Press Ltd.

Carr, C. T., Schrock, D. B., & Dauterman, P. (2012). Speech acts within Facebook status messages. *Journal Of Language & Social Psychology*, *31*, (176-196).

Celce-Murcia, Marianne, Brinton D. M. & Goodwin, J. M. (2010). *Teaching pronunciation*. New York, NY: Cambridge University Press.

Chafe, W. & J Danielewicz (1987), Properties of Spoken and Written Language. in R. Horowitz & S.J. Samuels, (Eds.), *Comprehending Oral and Written Language* (83-113). San Diego, CA: Academic Press Academic Press.

Chambers, J. K. (2002). Patterns of variation including change. In J.K Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change*. (349-372). Malden, MA: Blackwell Publishers Ltd.

Collins, B. & Mees, I. M.. (2003). *The phonetics of English and Dutch*. Köln, Germany: Die Deutsche Bibliothek

Coulmas, F. (1989). *The Writing Systems of the World*. Oxford, UK: Blackwell Publishers Ltd.

Coulmas, F. (2005). *Sociolinguistics: The study of speaker's choices*. Cambridge, UK: Cambridge University Press.

Coupland, N. (1980). Style-shifting in a Cardiff work-setting. *Language in Society, 9*, (1-12).

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge, UK: Cambridge University Press.

Cruttenden, A. (2001). *Gimson's pronunciation of English*. London, UK: Arnold.

Crystal, D. (1997). *A dictionary of linguistics and phonetics*. Cambridge, MA: Blackwell Publishers, Inc.

Crystal, D. (2001). *Language and the Internet.* Cambridge, UK: Cambridge University Press.

Crystal, D. (2005). *How language works*. New York, NY: The Overlook Press.

Crystal, D. (2008). *Txtng: The gr8 db8.* Oxford, UK: Oxford University Press.

Danet, B. & Herring, S. (2007). Introduction: Welcome to the multi-lingual Internet. In B. Danet & S. Herring (Eds.), *The multilingual Internet: Language, culture, and communication online.* (3-39). Oxford, UK: Oxford University Press.

De Fina, A., Schiffrin, D., & Bamberg, M. (2006). *Discourse and identity*. Cambridge, UK: Cambridge University Press.

Delahunty, G. P., and Garvey, J. J. (2010). *The English Language: From Sound to Sense*. Retrieved from http://wac.colostate.edu/books/sound/.

Ervin-Tripp, S. (2001). Variety, style-shifting, and ideology. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (1-18). Cambridge, UK: Cambridge University Press.

Edwards, J. (2009). *Language and identity*. Cambridge, UK: Cambridge University Press.

Fischer, J. L. (1958). Social influences in the choice of a linguistic variant. *Word*. 14 (47-56). In Labov 1972.

Fowler, J. (1986). *The social stratification of (r) in New York City department stores, 24 years after Labov*. Unpublished manuscript, New York University, New York City: New York. In Wardhaugh 2010.

Fouser, R., Inoue, N., & Lee, C. (2000). The pragmatics of orality in English, Japanese, and

Korean computer-mediated communication. In L. Pemberton & S. Shurville (Eds.), *Words on the web: Computer-mediated communication* (52-62). Exeter, UK: Intellect.

Gumperz, J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.

Hale, C. (1996). *Wired style: Principles of English usage in the digital age*. San Francisco, CA: HardWired.

Herring, S. (2001). Computer-mediated discourse. In D. Shiffrin, D. Tannen & H.E. Hamilton (Eds.), *The handbook of discourse analysis*. (613-634). Malden: Blackwell Publishers Ltd.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language.* Cambridge, UK:Cambridge University Press.

Kiesling, S. (1998). Men's identities and sociolinguistic variation: The case of fraternity men. *Journal of Sociolinguistics, 2*. (69-99).

Johnstone, B. (2010). Locating language in identity. In C. Llamas & D. Watt (Eds.), *Language and identities* (29-36). Edinburgh, UK: Edinburgh University Press Ltd.

Ko, K. (1996) Structural characteristics of computer-mediated language: A comparative analysis of InterChange discourse. *Electronic Journal of Communication, 6*. Retrieved from http://www.cios.org/www/ejc/v6n396.htm

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (2006). *The social stratification of English in New York City*. Cambridge, UK: Cambridge University Press.

Ling, R. (2005). The sociolinguistics of SMS: An analysis of SMS use by a random sample of Norwegians; In R. Ling & P. E. Pederson (Eds.), *Mobile communications: Renegotiation of the social space* (335-349). Kent, UK: Springer-Verlag London Limited.

Ly, B. (2012, August, 10). Qwerty keyboard wins the battle of the input methods. *Conversations by Nokia*. Retrieved from http://conversations.nokia.com/2012/08/10/qwerty-keyboard-wins-the-battle-of-the-input-methods/.

Mather, P. (2010). The social stratification of /r/ in New York City: Labov's department store study revisited. *Journal of English Linguistics, 40,* 338-356. Retrieved from http://ezproxy2.library.colostate.edu:4568/content/40/4/338.abstract

O'Connor, A. (2005). Instant messaging: Friend or foe of student writing. *New Horizons for Learning Online Journal 11.* Retrieved from http://newhorizons.org/strategies/literacy/oconnor.htm.

Palenchar, J. (2010). Keyboard, Touchscreen Phones Gain Q4 Share: NPD. *TWICE: This Week In Consumer Electronics*, *25*(5), 17.

Patrick P. L. (2002). The speech community. In J.K Chambers, P. Trudgill, and N. Schilling Estes (Eds.), *The handbook of language variation and change*. (573-598). Malden, MA: Blackwell Publishers.

Palfreyman, D. & Al Khalil, M. (200). "A funky language for teenzz to use": Representing Gulf Arabic in Instant Messenger. In B. Danet & S. Herring (Eds.), *The multilingual Internet: Language, culture, and communication online.* (43-63). Oxford, UK: Oxford University Press.

Rickford, J. R. & Eckert, P. (2001). Introduction. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (1-18). Cambridge, UK: Cambridge University Press.

Salkind, N. (2010). *Statistics for people who (think they) hate statistics*. New York, NY: Sage Publications, Inc.

Sampson, G. (1985). *Writing Systems*. Stanford, CA: Stanford University Press.

Shockey, L. (2003). *Sound patterns of spoken English*. Malden, MA: Blackwell Publishers Ltd.

Schreier, D. (2005). *Consonant change in English worldwide*. New York, NY: Pallgrave MacMillen.

Shuy, R. (1969). Sociolinguistic research at the Center for Applied Linguistics: the correlation of language and sex. *Giornata internazionale di sociolinguistica.* Rome: Palazzo Baldassini.

Squires, L. (2010). Enregistering internet language. *Language in Society, 39,* 457-492. Retrieved from http://ezproxy2.library.colostate.edu:3056/action/displayAbstract?fromPage=online&aid=7873122&fulltextType=RA&fileId=S0047404510000412

Tagliamonte, S., & Denis, D. (2008). Linguistic ruin? LOL instant messaging and teen language. *American Speech, 83* 3-34. Retrieved from http://ezproxy2.library.colostate.edu:2073/ehost/ pdfviewer/pdfviewer?sid=595dfc24-5625-4cbb-a868-59b4f412754c%40sessionmgr 115&vid=4&hid=117

fouser

Unger, J. M., & DeFrancis, J. (1995). Logographic and semasiographic writing systemd: A critique of Sampson's classification. *Neuropsychology and Cognition 7,* 45-48. Retrieved from http://link.springer.com/chapter/10.1007%2F978-94-011-1162-1_4#

Wardhaugh, R. (2010). *An introduction to sociolinguistics*. Oxford, UK: Blackwell Publishers Ltd.

Warschauer, M., Said, G., & Zohry, A. (2007). Language choice online: Globalization and identity. In B. Danet & S. Herring (Eds.), *The multilingual Internet: Language, culture, and communication online.* (303-318). Oxford, UK: Oxford University Press.

Appendix A

Demographic Information Survey

Thank you for volunteering to take part in the study. This is a preliminary survey form that you will need to complete **before** you transcribe your electronic messages. Please fill out the form completely and email it to the co-principal investigator at cahowe@rams.colostate.edu.

Please complete the following survey BEFORE you transcribe your messages.

1. How did you hear about this study?

2. Please indicate your level of education_____

    a. High School
    b. Bachelor's Degree
    c. Master's Degree
    d. PhD

3. What affiliation do you have with CSU?

    a. Current student
    b. Former student
    c. Faculty member
    d. Other

4. Please indicate your age_____

5. Do you have access to email, Facebook, and mobile phone text messages?

    Thank you again for your participation in the research study. Please email this completed form to cahowe@rams.colostate.edu.

Please remember that once you have completed the transcription sheet and/or forwarded your electronic messages to the Co-Principle Investigator (Chris Howe), you will be asked to complete one 10 minute follow-up questionnaire. Thank you!

Appendix B

<u>Instructions for submitting electronic text-based messages</u>:

The forms below are instructions for how to provide the co-principle investigator (Chris Howe) with electronic text-based messages. Please remember to you should only submit messages **you** created in the two weeks **prior** to receiving this form.

Your name will never be used in the research. Instead you will be assigned a screen name that will not give off any personal information about you.

Please keep in mind that you can submit any messages you wish, and you may refrain from submitting any messages you feel are too personal, private, etc. Please also keep in mind that the more messages you provide for research, the more valid the findings will be for future research.

Thank you again for all your participation.

Chris Howe

Please use the instructions below to send your messages to the investigator. Please note: you should submit messages you created in the two weeks **prior** to receiving this form.

Please use the codes below to identify the level of formality and the type of message. Write these codes after the end of each message you submit (there is an example on the next page).

1. (F) for formal
2. (I) for informal
3. (TM) for text message
4. (E) for email
5. (FB) for Facebook

Instructions for submitting Facebook messages:

1. You will be asked to submit Facebook posts you created in the two weeks **prior** to receiving this form.
2. Please only indicate that the message came from Facebook by typing (FB) after the end of the message.
3. Copy and paste the messages in one email and send it to the investigator at cahowe@rams.colostate.edu.

Instructions for submitting emails:

1. You will be asked to submit the emails you created in the two weeks **prior** to receiving this form.
2. Please use the appropriate codes above for each message. Type these codes after every message.
3. Copy and paste the messages in one email and send it to the investigator at cahowe@rams.colostate.edu.

Instructions for submitting mobile phone text messages:

1. You will be asked to submit the text messages you created in the two weeks **prior** to receiving this form.
2. Please use the appropriate codes above for each message. Type these codes after every message.
3. If your phone can forward text messages through email, please send all this information to the investigator at cahowe@rams.colostate.edu.

Or

4. If you cannot forward text messages through email, you can copy them in a document or in an email message. You can email this document or message to the investigator at cahowe@rams.colostate.edu. If you would like, you can use the form at the below to copy your messages and send them to the investigator.

Example email:

'Hi Fred-

I hope your weekend was nice. I just want to let you know about an event that will be going on this Friday. I attached the info in the doc.

Thanks
Julie **Codes for this email: E; F**

Example text message:
'Hmm. Def a possibility. Lemme run it by her n c' **Codes for this text message: TM; I**

Electronic Messages Transcription Form

Please complete the message transcription part with all the following information.

1. **Column 1: Transcribe the electronic message verbatim.**
   Please try to write the message as clearly and accurately as possible.

2. **Column 2: Please indicate the style of the message using the following codes.**

   (F) Formal
   (I) Informal

3. **Column 3: Please identify the type of message using the following codes.**

   (TM) Text Message
   (E) Email
   (FB) Facebook

Column 1	Column 2	Column 3

Appendix C

Examples of Most Frequent EMC Features in Proportion to Total Word Count.

Rank	Phonographic	%	Logographic	%	Semasiographic	%
1	!!!	0.35%	u	0.20%	☺	0.40%
2	…	0.31%	lol	0.05%	☹	0.05%
3	haha	0.30%	xoxo	0.03%	;)	0.01%
4	yeah	0.19%	w (with)	0.03%	:p	0.01%
5	oh	0.16%	tmro	0.03%	:-)	<0.01%
6	hey	0.15%	bday	0.02%	:/	<0.01%
7	ha	0.10%	ur	0.01%	:D	<0.01%
8	gonna	0.09%	omw	<0.01%	<3	<0.01%
9	ya	0.08%	bc	<0.01%	B-)	<0.01%
10	wanna	0.07%	r	<0.01%	>:(	<0.01%

Appendix D

Complete List of Features Found in the Corpus

Linguistic features	Formal Variant
Phonographic Feature	
Colloquial variants	
AGAIN	again
alritey	alright
ANY	any
arse	ass
AWESOME	awesome
aye	yes
BATTLE	battle
Be-atch	bitch
biggie	big (deal)
bikin	biking
bout	about
bro	brother
cause	because
cc'ing	courtesy copy(ing)
cell	cellular phone
checkin	checking
CHANGE	change
chillin	chilling
cmon	come on
cig	cigarette
cookin	cooking
combo	combination
comfy	comfortable
comin	coming
congrats	congratulations

convo	conversation
crocs	crocodiles
cus	cousin
cuz	because
da	the
dahhling	darling
def	definitely
docs	documents
der	there
doggies	dogs
doin	doing
dunno	don’t know
em	them
err	error
fam	family
fav	favorite
fellas	fellers
fer	for
flyer-ing	distributing flyers
fo	for
foco	Fort Collins
FOREVER	forever
freakin	freaking
freeezing	freezing
gimm(i)e	give me
goin	going
gonna	going to
goo	good
gotta	(have) got to
Gsa-ers	people who take part in the Graduate Student Association
gunna	going to
guuurl	girl

gving	giving
hafta	have to
hangin	hanging (out)
hey	hello
Hiiii	hi
holdin	holding
Holller	holler
horsey	horse
howdy	hello
howr	how are
Iffy	uncertainty
info	information
INTERNATIONALIZATION	internationalization
intro	introduction
ish	near or about
jumpin	jumping
jus	just
K	OK
kiddin	kidding
kiddos	kids
kinda	kind of
leavin	leaving
lemme	let me
lettin	letting
lil	little
lit	literature
lordy	lord
LOT	lot
lotta	lot of
LOVE	love
looove	love
luv	love

mac	Macintosh
makin	making
ME	me
min(s)	minute(s)
Mmkay	OK
n	and
Nah	no
NECTAR	nectar
NIGHT	night
NO	no
Nooooo	no
Noooooope	no
Nope	no
NOT	not
nothin	nothing
o	of
oi	hi
Okie Dokie	OK
ol	old
ONCE	once
ops	operations
Otay	OK
OUT	out
outta	out of
parkin	parking
pats	---
pepto	Pepto-Bismol
PERFECT	perfect
pics	pictures
pix	pictures
PLLEEEEAAAAASSSSS	please
pls	please

pooooooop	poop
pops	father
Prepping	preparing
PROBABLY	probably
prolly	probably
recs	requisites
round	around
RSVPing	sending out an RSVP
sayin	saying
sched	schedule
seein	seeing
sesh	session
sistah	sister
sitch	situation
sho	sure
SHOULD	should
shoulda	should have
smokin	smoking
sooo	so
soooooon	soon
sorta	sort of
spose	suppose
startin	starting
stooooreriiii	story
stat	immediately
sup	what's up?
sweeeet	sweet
ta	to
THANKIES	thanks
THANKS	thanks
tho	though
thru	through

thinkin	thinking
thx	thanks
til	until
till	until
TOMORROW	tomorrow
toooo	too
tryin	trying
TTTHHHAAAAANNNKKK	thank
txt	text
txting	texting
VERY	very
vibes	vibrations
vids	videos
wanna	want to
wantin	wanting
watcha	what are you
watchya	what are you
wassup	what’s up
well	we will
wer	we are
whaaaaat	what
Whaat	what
whatr	what are
whos	whose
wit	with
workin	working
ya	you
ya	yes
yea	yes
yeah	yes
yep	yes
yer	your

yeppers	yes
yoga-ers	people who take part in yoga
youuuuu	you
YYYOOOOUUUUUUUUU	you
yr	your
yup	yes
ze	the

Interjections

Aaaah	--
ahhhhh	--
aww/e	--
blah	--
booo	--
duh	--
eh	--
err	--
geez	--
gosh	--
golly	--
ha	--
hah	--
haha	--
hehe	--
hmm	--
huh	--
jaja	--
mmm	--
oooo	--
oh	--
Ooh	--
Ohh	--

ooops	--
phew	--
pleh	--
ugh	--
uh	--
uhm	--
um	--
umm	--
woohoo	--
wahoo	--
whammy	--
whew	--
yaaaaay	--
yay	--
YAYY	--
yeehaaa	--
yo	--
yooooo	--

Logographic Features

Acronyms

asap	as soon as possible
bbs	bulletin board system
bsg	--
btw	by the way
cc	courtesy copy
cpk	--
d	--
dt	downtown
ffxiv	final fantasy fourteen

fml	fuck my life
fyi	for your information
gd	good day
GTAship	Graduate Teaching Assistantship
hw	homework
lan	local area network
lol	laugh out loud
np	no problem
lyi	--
omg	oh my god
omw	on my way
oty	over the year
rbf	--
sb	--
ttyl	talk to you later
w	with
wtf	what the fuck

Abbreviations

abt	about
b-in-law	brother-in-law
bc	because
cali	California
camc	--
comp	computer
comps	computers
drs	doors
d-town	downtown
ebo	--
esp	especially
fri	Friday

fx	--
gp	--
grad	graduate
gvt	government
hr	human resources
hrs	hours
lab	laboratory
lang	language
lib	library
mon	Monday
morn	morning
msg	message
mtg	meeting
natl	natural
pcard	pin card
ppl	people
pres	presentation
prob	probably
prob	problem
prof	professor
rad	radical
rec	recreation
rep	repossession
res	restaurant
rev	review
sat	Saturday
spec	specification
sub	substitute
sun	Sunday
tdc	--
temp	temperature
thurs	Thursday

tmro	tomorrow
tmrw	tomorrow
tpk	--
tues	Tuesday
sry	sorry
uv	ultraviolet
vac	vacuum
wk	week
wed	Wednesday
vocab	vocabulary
xmas	Christmas

Rebus Writing

b	be
b4	before
c	see
r	are
u	you
u'd	you'd
u'll	you'll
ur	your
2	too
4	For

Typographic Symbols

xoxo	hugs and kisses
@	at

Semasiographic Features

:)	--

:(	--
:p	--
^^	--
:D	--
;)	--
:/	--
>:(	--
$$	--
;(*I*	--
:-)	--
B-)	--
:-*	--
<3	--